Institut d'études politiques de Paris

ECOLE DOCTORALE DE SCIENCES PO

Programme doctoral d'économie

Département d'économie

Doctorat en sciences économiques

THREE ESSAYS IN THE ECONOMICS OF INFORMATION

Daniel Martins de Almeida Barreto

Thesis supervised by Eduardo Perez-Richet, Associate Professor at Sciences Po, Department of Economics

Date of defense: May 23, 2023

Jury:

Jeanne Hagenbach (examiner – examinatrice)

Directrice de recherche CNRS, Département d'économie de Sciences Po (UMR 8259)

Emeric Henry

Professeur des Universités, Department of Economics, Sciences Po

Raphaël Lévy (referee – rapporteur)

Associate Professor, Department of Economics and Decision Sciences, HEC Paris

Eduardo Perez-Richet (supervisor – directeur de thèse)

Associate Professor, Department of Economics, Sciences Po

Vasiliki Skreta (examiner – examinatrice)

Professor, Department of Economics, University College London and Leroy G. Denman Regents Professor, Department of Economics, University of Texas at Austin

Nikhil Vellodi (examiner – examinateur)

Assistant Professor, Paris School of Economics

Adrien Henri Vigier (referee – rapporteur)

Professor of Economics, Faculty of Social Sciences, University of Nottingham

Acknowledgments

I am greatly indebted to many people that supported me academically and personally during this journey. These next paragraphs are an attempt at thanking some – although certainly not all – of them.

First, I would like to thank Eduardo Perez-Richet, who supervised me during this thesis. The process of learning how to do research is a hard one, where the path ahead often feels blurry, but I had the privilege of being able to count on him and to learn from his honesty, rigor and brilliancy. Also of great importance was Jeanne Hagenbach, who supported me since the beginning of (and even before) the program and was always available to listen to me and to guide me with her sharp and instructive feedback and her ever present good mood. I am also deeply thankful to Emeric Henry, whom I was lucky to be able to count on for feedback and advice and whom I deeply admire for his broad knowledge and generosity. I would also like to extend my gratitude to other members of Sciences Po's faculty, such as Franz Ostrizek, Junnan He, Michele Fioretti and Nicolas Coeurdacier, for all the support during the job market process.

At the core of this thesis are my coauthors, with whom I have had the privilege and pleasure of sharing the past years and who became much more to me than simple collaborators. I was extremely lucky to have had Victor on my cohort and to have developed with him a true partnership both intellectually and personally. Victor was throughout these years a great interlocutor for everything, a person I could talk with about anything from model misspecification or convex analysis to jazz or football or any personal issue that I would have. It was also a pleasure to get to know Alexis, whose curious mind helped to remind me of the joy of doing the work that we do. I will fondly remember all the afternoons that I spent with you both in front of a whiteboard trying to understand some model, even though I know that we will have many such afternoons in the future. Thank you both for the friendship, for all you taught me and for coping with my crazy conjectures and my poor jokes.

An important note is needed to thank my colleagues, the inhabitants of the combles with whom I shared most of my days. I am thankful for having received such

a warm welcome by Aseem, Charles, Elisa, Jan, Julia, Ludovic and Sophie, and to have shared the excitement of the early days (and the subsequent years) with Jérôme, Pauline and Pierre. Throughout the years the fauna of the combles became richer and more colorful, with the arrival of Aurélien, Ariane, Diego, Felipe, Gustave, Laurie, Mylène, Moritz, Naomi, Niklas, Rémi, Riddhi, Valentin and many others whom I had the pleasure of sharing so many lunches and coffee breaks. This paragraph can only be complete with a few more "thank you"'s. To Edgard, Ségal, Stefan and Victor le Grand for the laughs and the parties. To Juan and Leo for their infinite kindness, a sweet contrast to the bitterness of their always-present mate. To Sophia for her thoughtfulness. To Nicolò and Oliver for shifting from being my teachers to being my friends. To Marco for the complicity in sharing my passion for football. To Zydney for the dinners and the language coaching. To Olivia for the refreshing enthusiasm and curiosity. To Vladimir for the endless book recommendations that filled my "to read" list for the next thousand years. And finally to Sam, for a great friendship that I will take for the rest of my life and for having shown me his France (as well some opening ideas in the Caro-Kann).

While living in a foreign country can at times be tough, I am happy to have created in Paris a community that made me feel at home. I will always be grateful for having shared this chapter of my life with Ale, Amélie, Amanda, Andrew, Anna, Augustin, Flávia, Flore, Gedeão, Jules, Libertad, Mari, Paul, Pedro, Ros, Tristan... For having sung Brazilian songs late at night with Henrique and Bia. For the afternoons peeling clementines in the Seine with Liz. For having explored the hidden corners of this city with Luis. For the tenderness of Léa. For the relief of knowing that I can go visit Olivia and Hugo in Rio or Berlin or wherever they are. For the lightness and the laughs and the comfort of going out with Tony, Chepe, Rosa, Mulham, Habib and Khalil. For the incredible luck of having met, during my first weekend in Paris, the two people that would become my best friends in this city, Zoé and João.

Por último, me restam alguns agradecimentos a fazer em português. Ao Felipe, Capa, Bunduki, Zé, Luca, Léo e Tachi, pela amizade que sobrevive qualquer distância, temporal ou física. À Iná, pela presença constante em mim e pelo ímpar que somos nós. À minha família, e sobretudo aos meus pais, pelo apoio e amor incondicional que eu sempre recebi. Amo vocês.

Note to the reader. This thesis is composed of three independent papers written during my years as a PhD Candidate at Sciences Po Paris. The three papers are preceded by an introduction, which is intended as a historical contextualization of the field of Information Economics. A translation of the introduction in French is available after the English version. The bibliography for each of the papers is located at the end of each chapter, whereas the appendices are all located in a common section at the end of the thesis.

Contents

1 Price Discrimination with Redistributive Concerns **28**

1.1 Introduction 29

1.1.1 Related Literature 30

1.2 Model 32

1.2.1 Discussion of the Model 35

1.3 Three-Value Case 36

1.4 Optimal Segmentations 42

1.4.1 General Properties 42

1.4.2 Strongly Redistributive Social Preferences 43

1.4.3 Optimal Segmentations and Informational Rents 45

2 Persuading a Wishful Thinker **51**

2.1 Introduction 52

2.1.1 Related literature 54

2.2 Model 56

2.3 Receiver's wishful beliefs and behavior 59

2.4 Sender's value from persuasion 61

2.5 Applications 67

2.5.1 Information provision and preventive health care 67

2.5.2 Persuading a wishful investor 70

2.5.3 Public persuasion and political polarization 73

2.6 Conclusion 76

3 Text and Subtext **85**

3.1 Introduction 86

3.1.1 Related Literature 87

3.2 Model 88

3.2.1 Setup 88

3.2.2 Partial Understandings 89

3.2.3 Two Interpretations of the Model 89
3.3 Results 90
3.3.1 Feasibility 90
3.3.2 Optimality 91
3.4 Example 92

Appendices **99**

A Appendix for Chapter 1 **100**
A.1 Proof of lemma 1 100
A.2 Proof of lemma 2 100
A.3 Proof of proposition 2 101
A.4 Proof of proposition 3 103

B Appendix for Chapter 2 **105**
B.1 Proof of proposition 4 105
B.2 Overoptimism about preferred outcomes 106
B.3 Proof of lemma 3 107
B.4 Proof of proposition 5 111
B.5 Proof of proposition 8 113
B.6 Proof of proposition 7 115

C Appendix for Chapter 3 **116**
C.1 Proof of lemma 4 116
C.2 Proof of proposition 10 116

List of Figures

1.1 The Simplex representing M and two feasible segmentations. 37

1.2 Efficient Markets and Segmentations. 38

1.3 Uniformly Weighted Consumer-Optimal Segmentations. 39

1.4 Optimal Segmentations with Redistributive Preferences. 40

1.5 Rent Region. 41

1.6 Structure of optimal segmentations under strong redistributive preferences. 44

1.7 Segmentation σ^{NR}. 46

2.1 Comparison of supporting sets of beliefs. In blue, the set of Bayesian posteriors supporting action $a = 1$ for a Bayesian Receiver. In red, the set of Bayesian posteriors supporting action $a = 1$ for a wishful Receiver. 64

2.2 Expected payoffs under optimal information policies. Red curves: expected payoffs under wishful thinking. Blue curves: expected payoffs when Receiver is Bayesian. Dashed-dotted green lines: expected payoffs under a fully revealing experiment. 66

2.3 The Bayesian-optimal policy τ^B (in blue) vs. the wishful-optimal policy τ^W (in red) with respective supports $\{\mu^B_-, \mu^B_{0,2}\}$ and $\{\mu^W_-, \mu^W_{0,2}\}$. 67

2.4 The belief correspondence for $\varsigma = 2$, $c = 0.5$, $\alpha = 0.8$, $\underline{\theta} = 0.1$, $\overline{\theta} = 0.9$ and $\rho = 2$. Receiver is always overoptimistic concerning his health risk for any induced posterior, except at $\mu = 0$ or $\mu = 1$. Moreover, the belief threshold μ^W as a function of ρ is strictly increasing and admits μ^B as a lower bound. 69

2.5 Red (resp. blue) curves correspond to wishful (resp. Bayesian) Receiver. We set parameters to $c = 0.5$, $\alpha = 0.8$, $\underline{\theta} = 0.1$, $\overline{\theta} = 0.9$ and $\rho = 2$. Full lines correspond to the case where $\alpha = 1$ whereas dashed curves correspond to $\alpha = 0.8$. 70

2.6 Beliefs distortions in the electorate for $\rho = 2$, $\beta_1 = 1/4$, $\beta_2 = 1/2$ and $\beta_3 = 3/4$. Polarization equals $\pi(\mu) = 2(\eta(\mu, \beta^1) - \eta(\mu, \beta^3))$ which is maximized at $\mu^W(\beta^2) = 1/2$. 75

3.1 Three partitions satisfying a refinement order: In black P_1, in red P_2 and in blue P_3. 90
3.2 Sender's indirect utility 92
3.3 Recursive concavification 93
3.4 The value of persuasion under different modes of communication: text and subtext in Black, public in red and private in blue. 94

B.1 Functions α and μ^W for different payoff matrices $(u_a^\theta)_{a,\theta\in A\times\Theta}$. Action $a = 1$ is favored by a wishful Receiver whenever $\mu^W < \mu^B$. 110

Introduction

English version

Out of the many ways in which the 20th century might be remembered in the future, perhaps the most distinctive one is as the century in which the concept of information acquired a central status in the way human societies perceive the world and organize themselves. It was over that century that humanity discovered that all life forms rely on biological information stored in nucleic acids such as the DNA, or that uncertainty is a fundamental part of physical reality at the quantum scale. It was also over that century that the development of the personal computer and the internet revolutionized the way information is collected, processed and distributed, fundamentally changing how modern societies function. In the same way that thermodynamics and its concepts were the central categories in the epistemic regime prevalent during the industrial revolutions of the 18th and 19th centuries, information and its concepts became central epistemic categories during the digital revolutions of the 20th and 21st centuries.

The development of economic theory since the mid-20th century has echoed this trend. Work by Akerlof, Stigler, Stiglitz[1] and others rendered transparent how imperfect and asymmetrically distributed information play a central role in the determination of economic phenomena, demonstrating how previously analyzed equilibria could be fundamentally altered by even small changes in information. Information became not only one of the "fundamental particles" out of which microeconomic theories are built, alongside preferences, institutions and technology, but it also came to be regarded as a key commodity in any economy, sparking the beginning of a research agenda aimed at understanding how it is acquired, shared and used by economic agents.

This was not without technical challenges. Happily, language to talk about such concepts was already being developed. Notions related to the measurement and to

[1]see Akerlof (1970); Stigler (1961); Spence (1973); Stiglitz (1975); Rothschild and Stiglitz (1976); Milgrom and Roberts (1986).

ways of ordering information were developed by Shannon and Blackwell (Shannon, 1948; Blackwell, 1951), whereas equilibrium refinements accounting for incomplete information had been proposed in Harsanyi (1968).

Information Economics. It is worthwhile to briefly present some of the main strands of the literature in Information Economics, and how they relate to that broad agenda concerning the "life-cycle" of information.

The process of information acquisition is the subject of rational inattention models (Sims, 2003; Matějka and McKay, 2015), in which agents facing a certain decision problem choose what information to acquire given some cost[2]. Another literature exploring topics in information acquisition is the strategic experimentation literature (Bolton and Harris, 1999; Keller et al., 2005), in which agents need to choose how much of their resources (for instance their time) should be allocated on an uncertain alternative relative to another certain alternative. The strategic element comes into play when you consider that information from one agent might somehow flow to another agent, which modifies their incentives to explore the uncertain alternative and creates a free-rider problem.

The theme of information flowing between individuals is also present in the social learning literature (Banerjee, 1992; Smith and Sørensen, 2000), which studies how dispersed information is aggregated when agents can observe (and thus infer from) the actions taken by other agents. Typically information fails to be fully aggregated in such settings because agents might find it optimal to just take the same action as they see other agents taking, instead of conditioning their action on their own information.

Of course, one central way through which information flows is through communication. This is the focus of both the cheap talk (Crawford and Sobel, 1982) and the literature on verifiable message models (Grossman, 1981; Milgrom, 1981). Cheap talk models consider settings in which the agent that sends information is unconstrained, being able to choose any message costlessly (that is, being able to lie). Verifiable message models, on the other hand, study equilibria in situations in which the sender is able to choose how much of its information to transmit, but cannot lie.

Another relevant line of research concerns the way information is used by agents. While the bayesian paradigm provides a natural way to model the interpretation of evidence, experimental research has pointed out a number of ways in which people

[2]The cost of acquiring (or processing) some piece of information is typically considered to be proportional to the reduction in the Shannon entropy of the belief that it causes. Some limitations of the usage of these types of costs are discussed in Angeletos and Sastry (2019); Morris and Yang (2021); Nieuwerburgh and Veldkamp (2010).

might interpret evidence differently than what is considered in the bayesian model. Theoretical literature has mostly focused on studying how these different biases affect how information is translated into behavior, but has so far devoted less attention to another of its implications: how they shape incentives guiding how information is produced and shared. The two last chapters of this thesis tackle this issue, with each chapter considering a different deviation from the standard bayesian model.

A final strand of the literature that is important to mention is the one on Information Design (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019), which aims at understanding which informational environments are optimal under some objective, in different settings. The three papers contained in this thesis mainly relate to this literature. The first chapter, focused on a theme present in the digital economy, studies price discrimination through the lenses of data-driven market segmentation. It is concerned with market segmentations that are optimal for consumers and that prioritize poorer consumers. The second and third chapters are at the frontier of information design and behavioral economics: each of them explores the impact of a different deviation from the standard bayesian model into the design of information structures: the second chapter considers the impact of "wishful thinking" - the tendency of individuals of distorting their beliefs towards more optimistic scenarios, while the third chapter considers receivers with heterogeneous levels of understanding of the information conveyed. Below is a brief presentation of the themes present in each of the chapters.

Price Discrimination with Redistributive Concerns. Price discrimination is the subject of an extensive literature in Economics, dating back to Pigou (1920) and Robinson (1933). Historically, this literature would consider markets in which consumers were somehow exogenously segmented - for instance because they would be distributed geographically in a manner that would somehow reflect their characteristics -. Given some fixed segmentation of consumers, economists would try to infer conditions on the segmentation such that welfare (both producer's and consumer's) would be higher or lower relative to the case with an unsegmented market.

Recently, however, there has been a renewal in the interest devoted to this practice. This was prompted both by the increased practical relevance of this topic since the rise of digital markets, in which platforms possessing rich amounts of consumer data are able to flexibly segment consumers, as well as by developments in economic theory that made us more equipped to think analytically about this issue. Instead of thinking of consumer segmentations as exogenously given, this recent literature reasons at the space of all possible segmentations of consumers, allowing us to think

about what welfare outcomes are feasible in general (Bergemann et al., 2015) and to pin down segmentations that have a particular normative or positive appeal.

The aim of this paper is to study market segmentations aimed at benefitting consumers by lowering the prices they pay, and that prioritize poorer consumers in the sense that we are especially concerned by segmentations that will lower more the prices paid by poorer consumers. We show that while such redistributive segmentations are efficient (they maximize total surplus), they might not maximize aggregate consumer surplus. Instead, in the process of increasing the surplus of poorer consumers, some of the surplus that could potentially belong to some consumers ends up with the firm.

The results in this chapter characterize conditions on the aggregate composition of consumers such that this is true, and draws characteristics of redistributive segmentations.

Persuading a Wishful Thinker. The second chapter of this thesis is concerned with how biases on the receiving end of information affect incentives for information production and disclosure. We consider a model in which an interested sender devises an information structure to inform a biased receiver. The receiver is biased in that it distorts the informational content of the signal it observes, systematically holding beliefs that are more optimistic given its preferences.

We discuss the way in which such bias causes preferences to interact with beliefs and establish conditions for such biased receivers to be harder or easier to persuade. We use the insights from this model to illustrate why information campaigns might be ineffective at inducing preventive health behavior, how financial advisors might find it easier to sell riskier assets and how strategic information disclosure in elections might lead to increased polarization.

Text and Subtext. The third chapter, entitled "Text and Subtext" is devoted to analyzing information as a multi-layer concept. The basic idea explored in the chapter is that a piece of information might have varying degrees of depth, depending on the person interpreting it.

The idea of depth in a piece of information is one that is culturally familiar. Enlightenment philosophers were explicit about the distinction between the *exoteric* - the part of a text that was commonly understood - and *esoteric* - the aspects that could only be grasped by some - reading of philosophical texts, with authors such as Leibniz explicitly mentioning the deliberate usage of both modes as a strategy to make metaphysical writings acceptable to a more general (and, in his time, dogmatic) audience while still conveying the intended message to selected readers. A more recent

illustration of the strategic use of multi-layered information is in the phenomenon known as *dog whistling*: the usage, usually in political speeches, of coded language aimed at signaling something privately to some listeners without antagonizing others.

The aim of this chapter is to translate these ideas into the language of modern information design. We draw the joint distributions of beliefs that can be attained by any information structure when the audience varies in the depth that they can assess information, and draw a procedure that retrieves the value that a sender can obtain by exploiting such heterogeneity in understanding.

Bibliography

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Angeletos, G.-M. and Sastry, K. (2019). Inattentive economies. *NBER Working Paper*, (26413).

Banerjee, A. V. (1992). A Simple Model of Herd Behavior*. *The Quarterly Journal of Economics*, 107(3):797–817.

Bergemann, D., Brooks, B., and Morris, S. (2015). The limits of price discrimination. *American Economic Review*, 105(3):921–57.

Bergemann, D. and Morris, S. (2019). Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95.

Blackwell, D. (1951). Comparison of experiments. In Neyman, J., editor, *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, pages 93–102, Berkeley and Los Angeles. University of California Press. (Berkeley, CA, 31 July–12 August 1950). MR:0046002. Zbl:0044.14203.

Bolton, P. and Harris, C. (1999). Strategic experimentation. *Econometrica*, 67(2):349–374.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6):1431–1451.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law & Economics*, 24(3):461–483.

Harsanyi, J. C. (1968). Games with incomplete information played by "bayesian" players part ii. bayesian equilibrium points. *Management Science*, 14(5):320–334.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Keller, G., Rady, S., and Cripps, M. (2005). Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review*, 105(1):272–298.

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94(4):796–821.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391.

Morris, S. and Yang, M. (2021). Coordination and Continuous Stochastic Choice. *The Review of Economic Studies*, 89(5):2687–2722.

Nieuwerburgh, S. V. and Veldkamp, L. (2010). Information acquisition and under-diversification. *The Review of Economic Studies*, 77(2):779–805.

Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.

Robinson, J. (1933). *The Economics of Imperfect Competition*. London: Macmillan.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information*. *The Quarterly Journal of Economics*, 90(4):629–649.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Smith, L. and Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3):213–225.

Stiglitz, J. E. (1975). The theory of "screening," education, and the distribution of income. *The American Economic Review*, 65(3):283–300.

Version française

Parmi les nombreuses façons dont le XX^{e} siècle pourrait être considéré à l'avenir, la plus caractéristique est peut-être celle du siècle où le concept d'information a acquis une place centrale dans la manière dont les sociétés humaines perçoivent le monde et s'organisent. C'est au cours de ce siècle que l'humanité a découvert que toutes les formes de vie dépendent de l'information biologique stockée dans les acides nucléiques tels que l'ADN, ou que l'incertitude est une partie fondamentale de la réalité physique à l'échelle quantique. C'est également au cours de ce siècle que le développement de l'ordinateur personnel et de l'Internet a révolutionné la manière dont l'information est collectée, traitée et diffusée, changeant fondamentalement le fonctionnement des sociétés modernes. De la même manière que la thermodynamique et ses concepts ont été les catégories centrales du régime épistémique prévalent durant les révolutions industrielles des XVIII^{e} et XIX^{e} siècles, l'information et ses concepts sont devenus les catégories épistémiques centrales lors des révolutions numériques des XX^{e} et XXI^{e} siècles.

Le développement de la théorie économique depuis le milieu du XX^{e} siècle a suivi cette tendance. Les travaux d'Akerlof, Stigler, Stiglitz[3] et d'autres auteurs, ont mis en évidence le rôle central de l'information imparfaite et distribuée de manière asymétrique dans la détermination des phénomènes économiques. Ils ont démontré comment les équilibres précédemment analysés pouvaient être profondément modifiés par de légères variations de l'information. L'information est ainsi devenue l'un des éléments fondamentaux sur lesquels reposent les théories microéconomiques, aux côtés des préférences, des institutions et de la technologie. Elle a également été perçue comme un élément clé dans toute économie, donnant naissance à un programme de recherche visant à comprendre comment elle est acquise, partagée et exploitée par les acteurs économiques.

Cela n'a pas été sans soulever des défis techniques. Fort heureusement, un vocabulaire adapté à de tels concepts était déjà en cours d'élaboration. Les notions relatives à la quantification et à la comparaison des structures d'information ont été développées par Shannon et Blackwell (Shannon, 1948; Blackwell, 1951), alors que des concepts d'équilibre prenant en compte l'information incomplète ont été définis par Harsanyi (1968).

L'Économie de l'information. Il est utile de présenter brièvement certains des axes majeurs de la littérature en économie de l'information et leur lien avec le vaste

[3]voir Akerlof (1970); Stigler (1961); Spence (1973); Stiglitz (1975); Rothschild and Stiglitz (1976); Milgrom and Roberts (1986).

programme de recherche relatif au « cycle de vie » de l'information.

Le processus d'acquisition de l'information est l'objet des modèles d'inattention rationnelle (Sims, 2003; Matějka and McKay, 2015). Dans ceux-ci, les agents confrontés à un problème de décision choisissent quelles informations obtenir étant donné leur coût d'acquisition[4]. Un autre domaine de recherche lié à l'acquisition d'information concerne l'expérimentation stratégique (Bolton and Harris, 1999; Keller et al., 2005), où les agents doivent décider quelle proportion de leurs ressources (par exemple, leur temps) doit être allouée à une alternative incertaine plutôt qu'à une alternative certaine. L'aspect stratégique intervient lorsque l'on considère que l'information d'un agent peut être transmise d'une manière ou d'une autre à un autre agent, modifiant ainsi leurs incitations à explorer l'alternative incertaine et créant un problème de passager clandestin.

Naturellement, la communication est l'un des moyens fondamentaux par lesquels l'information se propage. Cela est au centre des modèles de communication sans coût (Crawford and Sobel, 1982) et de la littérature sur les modèles de communication certifiable (Grossman, 1981; Milgrom, 1981). Les modèles de communication sans coût considèrent des situations où l'agent qui envoie l'information n'est pas contraint et peut transmettre n'importe quel message (c'est-à-dire qu'il peut mentir). En revanche, les modèles de communication certifiable étudient les situations où l'expéditeur peut décider quelle partie de ses informations transmettre, mais ne peut pas mentir.

Un autre domaine de recherche pertinent concerne la manière dont les agents exploitent l'information. Bien que le paradigme bayésien propose une approche naturelle pour modéliser l'interprétation de l'information, la recherche expérimentale a identifié plusieurs façons dont les individus peuvent interpréter l'information différemment de ce qui est prévu dans le modèle bayésien. La littérature théorique s'est principalement concentrée sur l'étude de l'impact de ces divers biais sur la manière dont l'information se traduit en comportement, mais a accordé moins d'attention à une autre de ses implications : comment ces biais influencent les incitations qui guident la production et la diffusion de l'information. Les deux derniers chapitres de cette thèse traitent de cette problématique, chacun examinant une déviation différente par rapport au modèle bayésien standard.

Un dernier aspect de la littérature qu'il est important de mentionner est celui de la conception de l'information (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019), qui vise à comprendre quels environnements informationnels sont optimaux

[4]Le coût d'acquisition (ou de traitement) d'une information est généralement considéré comme proportionnel à la réduction de l'entropie de Shannon. Certaines limitations de l'utilisation de ces types de coûts sont discutées dans Angeletos and Sastry (2019); Morris and Yang (2021); Nieuwerburgh and Veldkamp (2010).

selon un objectif donné, dans différents contextes. Les trois articles contenus dans cette thèse se rattachent principalement à cette littérature. Le premier chapitre, axé sur un thème présent dans l'économie numérique, étudie la discrimination tarifaire à travers les prismes de la segmentation de marché basée sur les données. Il porte sur les segmentations de marché optimales pour les consommateurs et qui privilégient les consommateurs les plus pauvres. Les deuxième et troisième chapitres se situent à la frontière entre la conception de l'information et l'économie comportementale : chacun d'eux explore l'impact d'une déviation différente du modèle bayésien standard sur la conception des structures d'information : le deuxième chapitre examine l'impact des « croyances motivées » – la tendance qu'ont les individus à déformer leurs croyances dans la direction de leurs désirs – tandis que le troisième chapitre considère les destinataires ayant des niveaux de compréhension hétérogènes de l'information transmise. Voici une brève présentation des thèmes abordés dans chacun des chapitres.

Discrimination tarifaire et préoccupations redistributives. La discrimination tarifaire fait l'objet d'une vaste littérature en économie, remontant à Pigou (1920) et Robinson (1933). Historiquement, cette littérature considérait des marchés dans lesquels les consommateurs étaient segmentés de manière exogène, par exemple de manière géographique. Étant donnée une segmentation des consommateurs, les économistes ont cherché à déterminer les conditions dans lesquelles le bien-être (à la fois du producteur et du consommateur) est supérieur ou inférieur par rapport au cas d'un marché non segmenté.

Récemment, toutefois, l'intérêt porté à cette pratique a connu un regain. Cela a été dû à la fois à l'importance accrue de ce sujet dans le contexte des marchés numériques, où les plateformes disposant de riches données sur les consommateurs sont en mesure de segmenter les consommateurs de manière flexible, ainsi qu'aux développements de la théorie économique ayant permis de réfléchir de manière plus analytique à cette question. Au lieu de considérer les segmentations des consommateurs comme une donnée exogène, la littérature récente prend l'ensemble des segmentations possibles des consommateurs comme une variable de choix, ce qui nous permet de réfléchir aux résultats possibles en termes de bien-être en général (Bergemann et al., 2015) et d'identifier les segmentations ayant un attrait normatif ou positif particulier.

L'objectif de cet article est d'examiner les segmentations de marché qui visent à favoriser les consommateurs en diminuant les prix qu'ils paient, tout en accordant la priorité aux consommateurs les plus pauvres, étant donné que nous nous intéressons spécifiquement aux segmentations qui réduiront davantage les prix pour ces derniers. Nous montrons que si de telles segmentations redistributives sont efficientes au sens

de Pareto (elles maximisent le surplus total), elles peuvent ne pas maximiser le surplus des consommateurs. Au lieu de cela, dans le processus d'augmentation du surplus des consommateurs les plus pauvres, une partie du surplus qui pourrait potentiellement revenir à certains consommateurs se retrouve dans les mains de la firme.

Les résultats de ce chapitre caractérisent les conditions relatives à la composition globale des consommateurs pour lesquelles cela est vrai et mettent en évidence les caractéristiques des segmentations redistributives.

Persuasion et croyances motivées. Le deuxième chapitre de cette thèse examine comment les biais chez les destinataires de l'information influencent les incitations à produire et divulguer des informations. Nous étudions un modèle dans lequel un expéditeur conçoit une structure d'information afin de persuader un destinataire biaisé. Ce dernier est biaisé dans la mesure où il déforme le contenu informationnel du signal qu'il reçoit, en maintenant systématiquement des croyances allant dans le sens de ses préférences.

Nous analysons comment ce biais provoque une interaction entre les préférences et les croyances et déterminons les conditions dans lesquelles ces destinataires biaisés sont plus difficiles ou plus faciles à convaincre. Nous nous appuyons sur les enseignements de ce modèle pour illustrer pourquoi les campagnes d'information pourraient ne pas réussir à encourager un comportement préventif en matière de santé, comment les conseillers financiers pourraient trouver plus aisé de vendre des actifs plus risqués et comment la divulgation d'information stratégique lors des élections pourrait conduire à une polarisation accrue.

Texte et sous-texte. Le troisième chapitre, intitulé « Texte et sous-texte », se consacre à l'analyse de l'information en tant que concept à plusieurs niveaux. L'idée principale abordée dans ce chapitre est qu'une information peut présenter différents degrés de profondeur, selon la personne qui l'interprète.

La notion de profondeur de l'information est culturellement familière. Les philosophes des Lumières établissaient clairement la distinction entre la lecture *exotérique* – la partie d'un texte communément comprise – et *ésotérique* – les aspects accessibles seulement à certains – des textes philosophiques. Des auteurs tels que Leibniz mentionnaient explicitement l'usage délibéré des deux modes comme stratégie pour rendre les écrits métaphysiques acceptables pour un public plus large (et, à son époque, dogmatique), tout en transmettant le message voulu aux lecteurs choisis. Un exemple plus récent d'utilisation stratégique d'informations à plusieurs niveaux est le phénomène appelé « *dog whistling* » : l'emploi, généralement dans les discours

politiques, d'un langage codé ayant pour but de communiquer quelque chose en privé à certains auditeurs sans en froisser d'autres.

Le but de ce chapitre est de transposer ces idées dans le langage formel de la conception d'information. Nous établissons les distributions conjointes de croyances pouvant être atteintes par n'importe quelle structure d'information lorsque le public présente une diversité dans sa capacité à évaluer la profondeur de l'information. Nous proposons également une procédure permettant de déterminer le gain espéré qu'un émetteur peut obtenir en tirant parti d'une telle hétérogénéité de compréhension.

Bibliographie

Akerlof, G. A. (1970). The market for "lemons" : Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3) :488–500.

Angeletos, G.-M. and Sastry, K. (2019). Inattentive economies. *NBER Working Paper*, (26413).

Bergemann, D., Brooks, B., and Morris, S. (2015). The limits of price discrimination. *American Economic Review*, 105(3) :921–57.

Bergemann, D. and Morris, S. (2019). Information design : A unified perspective. *Journal of Economic Literature*, 57(1) :44–95.

Blackwell, D. (1951). Comparison of experiments. In Neyman, J., editor, *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, pages 93–102, Berkeley and Los Angeles. University of California Press. (Berkeley, CA, 31 July–12 August 1950). MR :0046002. Zbl :0044.14203.

Bolton, P. and Harris, C. (1999). Strategic experimentation. *Econometrica*, 67(2) :349–374.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6) :1431–1451.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law & Economics*, 24(3) :461–483.

Harsanyi, J. C. (1968). Games with incomplete information played by “bayesian” players part ii. bayesian equilibrium points. *Management Science*, 14(5) :320–334.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6) :2590–2615.

Keller, G., Rady, S., and Cripps, M. (2005). Strategic experimentation with exponential bandits. *Econometrica*, 73(1) :39–68.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices : A new foundation for the multinomial logit model. *The American Economic Review*, 105(1) :272–298.

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94(4) :796–821.

Milgrom, P. R. (1981). Good news and bad news : Representation theorems and applications. *The Bell Journal of Economics*, 12(2) :380–391.

Morris, S. and Yang, M. (2021). Coordination and Continuous Stochastic Choice. *The Review of Economic Studies*, 89(5) :2687–2722.

Nieuwerburgh, S. V. and Veldkamp, L. (2010). Information acquisition and under-diversification. *The Review of Economic Studies*, 77(2) :779–805.

Pigou, A. C. (1920). *The Economics of Welfare.* London : Macmillan.

Robinson, J. (1933). *The Economics of Imperfect Competition.* London : Macmillan.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets : An Essay on the Economics of Imperfect Information*. *The Quarterly Journal of Economics*, 90(4) :629–649.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27 :379–423.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3) :665–690.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3) :355–374.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3) :213–225.

Stiglitz, J. E. (1975). The theory of "screening," education, and the distribution of income. *The American Economic Review*, 65(3) :283–300.

Chapter 1

Price Discrimination with Redistributive Concerns

Abstract

Consumer data can be used to sort consumers into different market segments, allowing a monopolist to charge different prices at each segment. We study consumer-optimal segmentations with redistributive concerns, i.e., that prioritize poorer consumers. Such segmentations are efficient but may grant additional profits to the monopolist, compared to consumer-optimal segmentations with no redistributive concerns. We characterize the markets for which this is the case and provide a procedure for constructing optimal segmentations given a strong redistributive motive. For the remaining markets, we show that the optimal segmentation is surprisingly simple: it generates one segment with a discount price and one segment with the same price that would be charged if there were no segmentation.

[0]This chapter is joint work with Alexis Ghersengorin and Victor Augias. We thank Eduardo Perez-Richet for his guidance on this project. We also thank Matthew Elliott, Jeanne Hagenbach, Emeric Henry, Emir Kamenica, Frédéric Koessler, Shengwu Li, Franz Ostrizek, Nikhil Vellodi, Colin Stewart, and seminar participants at Sciences Po, Paris School of Economics, Northwestern University, University of Konstanz, CUNEF, University of Rome "Tor Vergata", University of Barcelona, University of Amsterdam and WU Vienna for helpful discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 850996 – MOREV and 101001694 – IMEDMC)

1.1 Introduction

Consumers are continuously leaving traces of their identities on the internet, be it through social media activity, search-engine utilization, online-purchasing and so on. The vast amount of consumer data that is generated and collected has acquired the status of a highly-valued good, as it allows firms to tailor advertisements and prices to different consumers. In practice, the availability of consumer data *segments* consumers: observing that a given consumer has certain characteristics allows firms to fine-tune how they interact with people that share those characteristics. Adjusting how coarse-grained the information available about consumers is impacts how they will be segmented, what sort of digital market interactions they will have and what prices they will pay. This suggests room for regulatory oversight.

As shown by Bergemann et al. (2015), consumer segmentation and price discrimination can induce a wide range of welfare outcomes. It can not only be used to increase social surplus—by creating segments with prices that allow more consumers to buy—, but can also be performed in a way that ensures that all created surplus accrues to consumers — that is, that maximizes consumer surplus. This is done by creating segments that pool together consumers with high and low willingness to pay, thus allowing higher willingness to pay consumers to benefit from lower prices. However, an important aspect of price discrimination that remains overlooked by the literature is its *distributive effect*: since different consumers pay different prices, this practice defines how surplus is distributed *across* consumers, raising questions about how it can benefit poorer consumers relative to richer ones. Indeed, if willingness to pay and wealth are positively related, segmentations that maximize total consumer surplus tend to benefit richer consumers.

In this paper we provide a normative analysis of the distributive impacts of market segmentation. Our aim is to study how this practice impacts different consumers and how it should be performed under the objective of increasing consumer welfare while prioritizing poorer consumers. Our results draw qualitative characteristics of segmentations that achieve this goal, which can be used to inform future regulation. Importantly, our analysis also shows that the prioritization of poorer consumers can be inconsistent with the maximization of total consumer surplus: raising the surplus of poorer consumers may only be possible while granting additional profits to the producer, at the expense of richer consumers.

We consider a setting in which a monopolist sells a good on a market composed of heterogeneous consumers, each of whom can consume at most one unit and is characterized by their willingness to pay for the good. A social planner can provide

information about consumers' willingness to pay to the monopolist. The information provision strategy effectively divides the aggregate pool of consumers into different *segments*, each of which can be priced differently by the monopolist. The social planner's objective is to maximize a weighted sum of consumers' surplus. As in Dworczak et al. (2021), we consider weights that are decreasing on the consumer's willingness to pay, capturing the notion of a redistributive motive under the assumption that consumers with higher willingness to pay are on average richer than those with lower willigness to pay.

We first establish that optimal segmentations are Pareto efficient, such that satisfying a redistributive objective does not come at the expense of social surplus. Bergemann et al. (2015) show that, in the absence of redistributive concerns, consumer-optimal segmentations do not strictly benefit the monopolist: all of the surplus created by the segmentation accrues to consumers. In contrast, we show that once redistributive preferences are considered, consumer-optimal segmentations may imply additional profits to the monopolist. This happens because increasing the surplus of poor consumers is done by pooling them with even poorer consumers, such that they can benefit from lower prices. In doing so, richer consumers become more representative in other segments, which might increase the price they pay. We characterize the set of markets for which this is the case and denote them as rent markets. For no-rent markets, on the contrary, *any* redistributive objective can be met while still maximizing total consumer surplus. In this case, our analysis selects one among the many consumer-optimal segmentations established by Bergemann et al. (2015). These insights are illustrated through a three-type example in section 1.3.

Our analysis also provides insights on how to construct optimal segmentations. We show that, in no-rent markets, consumer-optimal segmentations with redistributive concerns exhibit a stunningly simple form, simply dividing consumers into two segments: one where the price is the same that would be charged under no segmentation and one with a discount price. In rent markets, we show that consumer-optimal segmentations under sufficiently strong redistributive preferences divide consumers into contiguous segments based on their willingness to pay, having consumers with the same willingness to pay belong to at most two different segments. This allows us to construct a procedure that generates consumer-optimal segmentations under strong redistributive preferences, which is discussed in section 1.4.2.

1.1.1 Related Literature

Third-degree price discrimination and its welfare effects are the subject of an extensive literature. Early analysis (Pigou, 1920; Robinson, 1933) and subsequent development

(Schmalensee, 1981; Varian, 1985) considered exogenously fixed market segmentations and studied conditions under which such segmentations would increase or decrease total surplus.

This literature has recently undergone a transformation, prompted by both technical innovations in microeconomic theory and the change in character of the practice of price discrimination brought about by the ascent of digital markets. Recent developments incorporate an information design approach to study the welfare impacts of third-degree price discrimination over *all possible* market segmentations, rather than taking a segmentation as exogenously fixed. Bergemann et al. (2015) analyze a setting with a monopolist selling a single good and characterize attainable pairs of consumer and producer surplus, showing that any distribution of total surplus over consumers and producer that guarantee at least the uniform-price profit for the producer is attainable. In particular, they show that there are typically many consumer-optimal segmentations of a given market. Their analysis has been extended to multi-product settings by Haghpanah and Siegel (2022a,b) and to imperfect competition settings by Elliott et al. (2021) and Ali et al. (2022). Hidir and Vellodi (2020) study market segmentation in a setting where the monopolist can offer one from a continuum of goods to each consumer, such that consumers, upon disclosing their information, face a trade-off between being offered their best option and having to pay a fine-tuned price. Finally, Roesler and Szentes (2017) and Ravid et al. (2022) study the inverse problem of information design to a buyer who is uncertain about the value of a good. Our paper differs from these by focusing on how surplus is distributed *across* consumers, and by studying consumer-optimal segmentations when different consumers are assigned different welfare weights. We show that, once distributional preferences are taken into account, optimal segmentations might not coincide with consumer-optimal segmentations under uniform welfare weights. When they do, our analysis selects one among the many direct consumer-optimal segmentations established in Bergemann et al. (2015).

Our paper also dialogues with a recent literature on mechanism design and redistribution, most notably with Dworczak et al. (2021) and Akbarpour et al. (2020), who study the design of allocation mechanisms under redistributive concerns; and Pai and Strack (2022), who study the optimal taxation of a good with a negative externality when agents differ on their utility for the good, disutility for the externality and marginal value for money. A key difference in the results obtained in these papers and ours is that, in their settings, redistributive mechanisms are not pareto-efficient: redistribution implies some loss in social surplus. This is not the case in our paper, where optimal redistributive segmentations always maximize total surplus.

Finally, our paper dialogues with Dube and Misra (2022), who study experimentally the welfare implications of personalized pricing implemented through machine learning. The authors find a negative impact of personalized pricing on total consumer surplus, but note that a majority of consumers benefit from price reductions under personalization, pointing that under some inequality-averse weighted welfare functions, data-enabled price personalization might increase welfare. Their paper shows experimentally how the implementation of market segmentations aimed at maximizing profits might generate, as a by-product, the redistribution of surplus among consumers. Our paper, on the other hand, shows theoretically how consumer-optimal redistributive segmentations might grant additional profits for the firm.

1.2 Model

A monopolist (he) sells a good to a continuum of mass one of buyers, each of whom can consume at most one unit. We normalize the marginal cost of production of the good to zero. The consumers privately observe their type v, which corresponds to their willingness to pay for the good. We assume that the consumers' type can take a finite number K of possible values $V = \{v_1, \dots, v_K\}$, where $0 < v_1 < \dots < v_K$. We let $\mathcal{K} := \{1, \dots, K\}$. A *market* μ is a distribution over the valuations. We denote the set of all possible markets:

$$M := \Delta(V) = \Big\{\mu \in \mathbb{R}^K \mid \sum_{k \in \mathcal{K}} \mu_k = 1 \text{ and } \mu_k \geq 0 \text{ for all } k \in \mathcal{K}\Big\}.$$

Price v_k is *optimal for market* $\mu \in M$ if it maximizes the expected revenue of the monopolist when facing market μ, that is:

$$v_k \sum_{i=k}^{K} \mu_i \geq v_j \sum_{i=j}^{K} \mu_i, \quad \forall j \in \mathcal{K}.$$

Let M_k denote the set of markets where price v_k is optimal. It is given by:

$$M_k = \Big\{\mu \in M \,\Big|\, v_k \in \arg\max_{v_i \in V} v_i \sum_{j=i}^{K} \mu_j\Big\},$$

for any $k \in \mathcal{K}$. In the remaining of the paper we will hold an aggregate market fixed and denote it by $\mu^0 \in M$.

Segmentation. The consumers' types are perfectly observed by a social planner (she) who can *segment* consumers, that is, sort consumers into different sub-markets. The set of possible segmentations of an aggregate market μ^0 is given by:

$$\Sigma(\mu^0) := \left\{ \sigma \in \Delta(M) \,\middle|\, \int_{\Delta(M)} \mu \, \sigma(\mathrm{d}\mu) = \mu^0 \right\}.$$

Formally, a segmentation is a probability distribution on M which averages to the aggregate market μ^0. The requirement that the different segments generated by a segmentation average to the aggregate market ensures that the segmentation simply sorts existing consumers into different groups, without fundamentally altering the aggregate composition of consumers in a market. This requirement is akin to the Bayes Plausibility condition that is typically used in the Bayesian Persuasion literature (Kamenica and Gentzkow, 2011).

Given a segmentation σ, the monopolist can price differently at each segment μ in the support of σ. A pricing rule is a mapping $p\colon M \to V$. As will become clear in problem 1.4, segments with more than one optimal price play a key role in our results. We focus on the following pricing rule:

$$p(\mu) = \min \left\{ \underset{k \in \mathcal{K}}{\arg\max}\, v_k \sum_{i=k}^{K} \mu_i \right\}.$$

At each segment, the monopolist charges the smallest price among all optimal prices in that segment. This pricing rule makes the objective of the social planner (stated in equation (P)) upper semi-continuous and ensures the existence of an optimal segmentation[1].

Social objective. The social planner's objective is to maximize a weighted sum of consumers' surplus, with positive weights $\lambda \in \mathbb{R}_+^K$. Each dimension λ_k of the vector λ corresponds to the marginal contribution to social welfare of consumers of type v_k. The surplus of a consumer of type v_k in market μ is given by:

$$U_k(\mu) := \max \{0, v_k - p(\mu)\}.$$

[1]Although technically important, this pricing rule does not impact our results qualitatively. Indeed, any joint distribution of consumers and prices that can be induced by the social planner under this pricing rule could be approximated arbitrarily well by a social planner facing a monopolist who selects among optimal prices in some other way.

The weighted consumer surplus on market μ is given by:

$$W(\mu) \coloneqq \sum_{k\in\mathcal{K}} \lambda_k \, \mu_k \, U_k(\mu),$$

for any $\mu \in M$. Hence, for any aggregate market μ^0, the social planner's objective is given by the following maximization program:

$$\max_{\sigma\in\Sigma(\mu^0)} \int_{\Delta(M)} W(\mu)\, \sigma(\mathrm{d}\mu). \tag{P}$$

Given an aggregate market μ^0, a segmentation $\sigma \in \Sigma(\mu^0)$ is *optimal* if it solves (P). We focus on welfare weights that are decreasing on the consumer's willingness to pay, such that $\lambda_k \geq \lambda_{k'}$ for any $k < k' \leq K-1$, and say that the social planner has *redistributive preferences* if the inequality holds strictly for some $k, k' \in \mathcal{K}$. Under the assumption that consumers with lower willingness to pay are on average poorer than consumers with higher willingness to pay, this amounts to attributing a greater weight to surplus accruing to poorer consumers[2].

Efficiency. Every consumer has a value for the good that is strictly greater than the marginal cost of production. Hence, social surplus is maximized when every consumer buys the good. We say that a market μ is *efficient* if every consumer can buy the good, that is, if the lowest optimal price for the seller at that market allows everyone to consume: $p(\mu) = \min \operatorname{supp}(\mu)$. For a given market μ and Pareto weights λ, the maximum feasible social surplus is thus given by

$$s(\mu) = \sum_{k\in\mathcal{K}} \lambda_k \mu_k v_k.$$

Note that a segmentation of μ achieves $s(\mu)$ if and only if it is efficient. A segmentation σ is *efficient* if it is only supported on efficient markets.

Informational Rents. The profit of the monopolist at market μ is given by:

$$\pi(\mu) = p(\mu) \sum_{k\in C_{p(\mu)}} \mu_k,$$

[2]We follow here the approach by Dworczak et al. (2021).

where $C_p = \{k \in \mathcal{K} \mid v_k \geq p\}$. The profit of the monopolist under segmentation σ is given by:

$$\Pi(\sigma) = \int_{\Delta(M)} \pi(\mu)\, \sigma(\mathrm{d}\mu)$$

Segmenting the aggregate market can only weakly increase the expected profit of the monopolist relative to no segmentation. Therefore, we always have $\Pi(\sigma) \geq \pi(\mu^0)$ for any $\sigma \in \Sigma(\mu^0)$. We say that some segmentation σ grants a *rent* to the monopolist whenever $\Pi(\sigma) > \pi(\mu^0)$.

Uniformly Weighted Consumer-Optimal Segmentations. If $\lambda_k = \lambda_{k'} > 0$ for all $k, k' \in \mathcal{K}$, program (P) corresponds to the maximization of the total consumer surplus over all possible segmentations. A segmentation that solves this optimization problem is named *uniformly weighted consumer-optimal.* As shown in Bergemann et al. (2015), uniformly weighted consumer-optimal segmentations are (i) efficient—and hence achieve the maximum feasible social surplus—, and (ii) do not grant the monopolist any rent. For an interior aggregate market μ^0, there exists infinitely many uniformly weighted consumer-optimal segmentations. In section 1.4.3, we characterize the set of aggregate markets for which consumer-optimal segmentations *with redistributive preferences* are also uniformly weighted consumer-optimal, thus providing a natural way to select among these segmentations for such markets.

1.2.1 Discussion of the Model

Information Provision as Segmentation. In digital markets, information provision about consumers often occurs through the assignment of *labels* to different consumers. Indeed, one could think of a model in which the social planner adopts a signal structure $\ell\colon V \to \Delta(L)$, where L is a set of labels. The meaning of each label is then pinned down by the social planner's strategy, and the monopolist optimally chooses different prices for consumers with different labels.

Such a model is equivalent to ours. Indeed, any segmentation $\sigma \in \Sigma(\mu^0)$ can be implemented by some signal structure ℓ, and any signal structure ℓ implements some segmentation $\sigma \in \Sigma(\mu^0)$. The approach of working directly in the space of feasible distributions over markets rather than in the space of labeling strategies is standard in the information design literature (Kamenica and Gentzkow, 2011).

Continuum of Consumers. While we consider a setting with a continuum of consumers, our model is equivalent to one in which there is a discrete number of consumers, with types independently distributed according to μ^0. Under this

interpretation, the social planner commits ex-ante to an information structure σ to inform the monopolist, which defines the distribution of posterior beliefs μ that the monopolist will form upon facing each consumer.

1.3 Three-Value Case

In this section, we illustrate our model and some of the results from the following sections in the simple three-value case.

Setup. Let's consider three types, $v_1 = 1$, $v_2 = 2$ and $v_3 = 3$. We can conveniently depict the set of markets M as the two-dimensional unit simplex (see Mas-Colell et al., 1995, p.169). It is depicted in figure 1.1, where each vertex of the simplex represents a degenerate market on a value $v \in V$, denoted by the Dirac measure δ_v.

In the left panel of figure 1.1 are drawn the three different price regions M_1, M_2 and M_3. The points in each of the regions correspond to the markets for which each of the different prices $\{1, 2, 3\}$ are optimal for the monopolist[3]. The border between two adjacent regions represents markets for which there are more than one optimal price. Given pricing rule p, the price charged in such markets is the lowest amongst the optimal.

In the right panel, an aggregate market $\mu^0 = (0.3, 0.4, 0.3)$ is represented, which is in the interior of the region M_2, meaning that v_2 is a strictly optimal price for μ^0. Two possible segmentations are depicted: the one in green dashed lines, that segments μ^0 into the three degenerate markets (thus implementing first-degree price discrimination); and the one in black dotted lines, that segments μ^0 into three segments: μ', containing types all three types and being priced v_1; μ'', containing only types v_2 and v_3 and being priced v_2; and μ''', containing all three types and being priced v_3.

Any splitting of μ^0 into a set of points $S \subset M$ represents a feasible segmentation, as long as $\mu^0 \in \text{co}(S)$[4]. A segmentation is optimal given weights $(\lambda_1, \lambda_2, \lambda_3)$, with $\lambda_1 \geq \lambda_2 \geq \lambda_3$, if it maximizes the sum of weighted consumer surplus over all segments generated. Note that consumers of type v_1 never get any consumer surplus (since the monopolist never charges a price lower than their willingness to pay), such that the optimal segmentation trades-off surplus obtained by types v_2 and v_3. We will focus, without loss of generality, on direct segmentations, i.e. segmentations in which there is not more than one segment with a given price.

[3]Formally, for any k, $M_k = \text{cl}(p^{-1}(v_k))$, where $\text{cl}(S)$ denotes the topological closure of a generic set S.

[4]For any set S, $\text{co}(S)$ denotes the convex hull of S

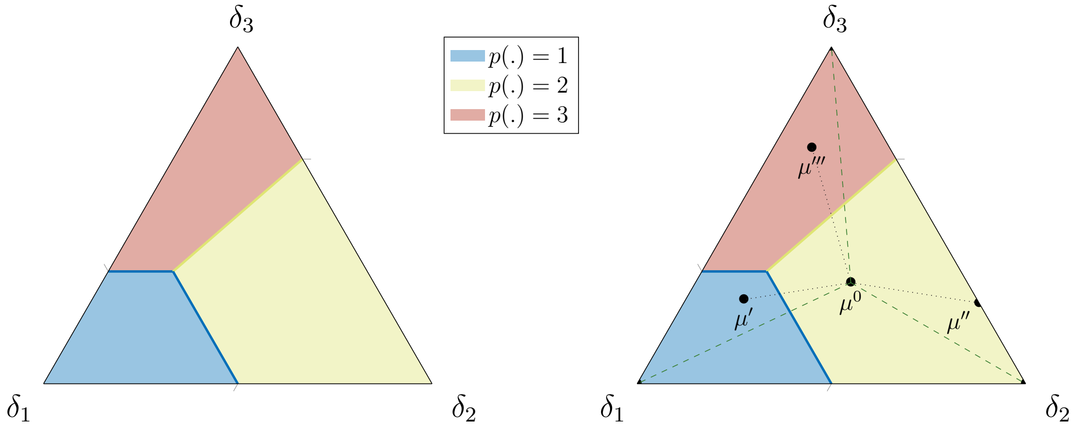


Figure 1.1: The Simplex representing M and two feasible segmentations.

General Properties of Optimal Segmentations. A first step for finding the optimal segmentation of μ^0 is to observe that any optimal segmentation must be efficient. To see that, consider the black dotted segmentation in the right panel of figure 1.1. Both μ' and μ'' are efficient, since all the consumers in these segments are able to buy the good. The remaining segment μ''', however, is not efficient, as it contains some consumers with type v_1 and v_2 who are not able to consume under that segment's price. One could solve that by re-segmenting μ''' in the following way: creating a segment μ_b''' containing all of the types v_1 and v_2 and some of the types v_3 that used to belong to μ''', and another segment δ_3 containing only the remaining types v_3. Note that the amount of type v_3 in μ_b''' can be adjusted to ensure that this segment will have price v_1. That way, both of the resulting segments will be efficient. Furthermore, this re-segmentation of μ''' *unambiguously* increases consumer welfare, since it has no impact on the welfare of consumers in μ' and μ'' and (weakly) increases the surplus of every consumer previously belonging to μ'''.

Indeed, a welfare-increasing segmentation can be performed to any inefficient market. This narrows down the search for an optimal segmentation, as we know that it must be supported *only* on efficient segments. The left panel of figure 1.2 depicts, in orange, the efficient markets. These are: the degenerate market δ_3; the set of markets in region M_2 that have no consumer with value 1; and the entire region M_1.

We can further note that, in an optimal segmentation, the segment with price v_1 must not belong to the interior of region M_1. To see that, consider the right panel of figure 1.2. In it are depicted two segmentations: σ_a, which splits μ^0 into μ_a and μ', and σ_b, which splits μ^0 into μ_b and μ'. Segmentation σ_b is always preferred over σ_a for two reasons. First, μ_b has a higher share of types v_2 and v_3 than μ_a. Since these are the only two types that are extracting surplus on the segment whose price is v_1, having a higher share of them increases the social planner's objective. Second, μ_b is

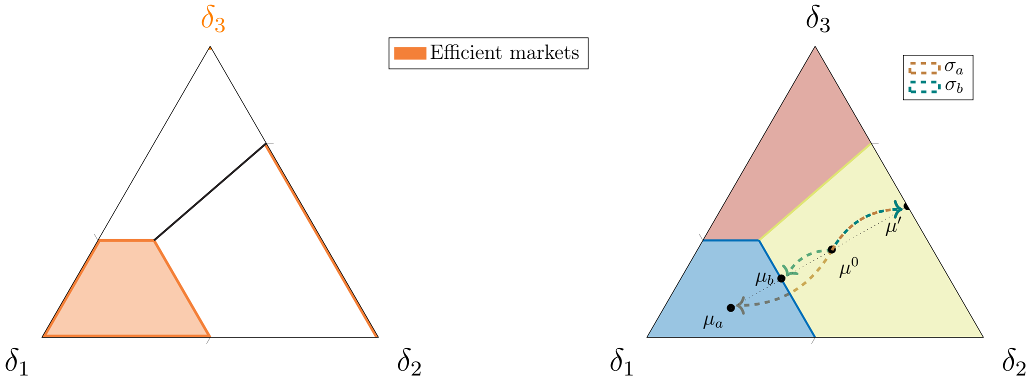


Figure 1.2: Efficient Markets and Segmentations.

"closer" to μ^0, which means that $\sigma_b(\mu_b) > \sigma_a(\mu_a)$. That means that segmentation σ_b is able to include a bigger mass of consumers in the segment where they will extract the largest surplus, thus also increasing the social planner's objective.

The argument outlined above illustrates how every segmentation generating a segment on the interior of region M_1 must be dominated by some segmentation that instead generates a segment on the boundary of regions M_1 and M_2. This amounts to saying that any optimal segmentation must include a segment in which the monopolist is indifferent between charging price v_1 or charging some other price. The intuition for that is simple: if the monopolist strictly prefers to charge price v_1 in that segment, then there's still room for "fitting" other types in that segment in a Pareto improving way.

Uniformly Weighted Consumer-Optimal Segmentations. We begin by considering the case where $\lambda_1 = \lambda_2 = \lambda_3$. The left panel of figure 1.3 depicts three different segmentations, σ_a, σ_b and σ_c, each of them generating one segment with price v_1 and one segment with price v_2. All of these three segmentations are uniformly weighted consumer-optimal. This follows from the fact that i) they maximize total (consumer + producer) surplus, since they are all efficient, and ii) the monopolist does not get any of the surplus that is created from the segmentation [5].

Indeed, there are uncountably many uniformly weighted consumer-optimal segmentations of μ^0. All of these are equivalent in that they maximize total consumer surplus, but they are not equivalent in how they distribute such surplus *across*

[5]One way of seeing this is as follows: A decision-maker strictly benefits from observing a piece of information if, as a result of this observation, she is able to make better decisions than she would have made absent this information. In our setting, this amounts to the monopolist being able to, as a result of the segmentation, choose *different* prices than the uniform price, at markets in which these different prices are *strictly* preferred over the uniform price. Since price v_2 belongs to the set of optimal prices in every segment generated by the segmentations in figure 1.3, the monopolist does not strictly benefit from them.

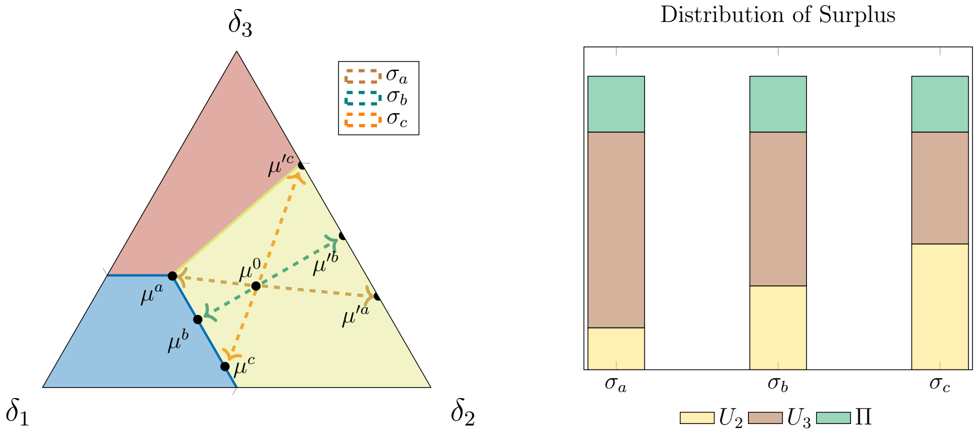


Figure 1.3: Uniformly Weighted Consumer-Optimal Segmentations.

consumers. This can be seen in the right panel of figure 1.3: while the three segmentations of the left panel induce the same profit for the monopolist and the same total consumer surplus, σ_c induces greater surplus for consumers of type v_2 than the other segmentations. This is so because, among the segments priced at v_1, μ_c is the one that includes the most consumers of type v_2, who can then benefit from a low price.

Consumer-Optimal Segmentations under Redistributive Preferences. Let's now consider the case when $\lambda_2 > \lambda_3$. Among the segmentations depicted in the left panel of figure 1.3, segmentation σ_c is now preferred over σ_a and σ_b. But is it optimal? One way of increasing the surplus of consumers of type v_2 further is to exchange consumers between the two segments generated by σ_c: by exchanging the remaining consumers of type v_3 that are present in μ^c against some of the consumers of type v_2 present in μ'^c, one can increase the amount of types v_2 that pay a low price. While this exchange increases the surplus of types v_2, it dramatically decreases the surplus of types v_3, since now there are sufficiently many of them in segment μ'^c for the monopolist to want to increase the price charged at that segment. This would lead to a segmentation that is no longer uniformly weighted consumer-optimal: the price increase in segment μ'^c would cause some of the surplus that was previously captured by consumers of type v_3 to now be granted to the monopolist instead. The result below establishes when this exchange is desirable from the social planner's perspective.

Result 1. *Let $\mu^0 = (0.3, 0.4, 0.3)$. Then, the two following assertions are satisfied:*

(i) If the inequality

$$\frac{\lambda_2}{\lambda_3} < \frac{v_3 + v_2 - v_1}{v_2 - v_1},$$

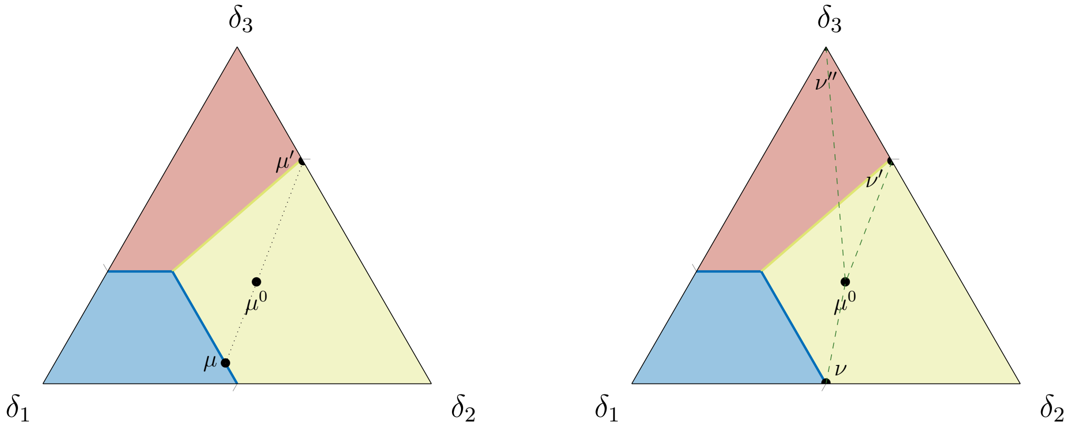


Figure 1.4: Optimal Segmentations with Redistributive Preferences.

is satisfied, then the consumer-optimal segmentation under redistributive preferences is also uniformly weighted consumer-optimal and generates two segments. One supported on $\{v_1, v_2, v_3\}$ *and the other one supported on* $\{v_2, v_3\}$*. This segmentation is represented in the left panel of figure 1.4;*

(ii) If the inequality

$$\frac{\lambda_2}{\lambda_3} > \frac{v_3 + v_2 - v_1}{v_2 - v_1},$$

is satisfied, then the consumer-optimal segmentation under redistributive preferences is not *uniformly weighted consumer-optimal and generates three segments. The first one is supported on* $\{v_1, v_2\}$*, the second is supported on* $\{v_2, v_3\}$*, and the third is supported on* $\{v_3\}$*. This segmentation is represented in the right panel of figure 1.4.*

An important consequence of this result is that if the social planner's preferences are sufficiently redistributive, meaning that λ_2 is sufficiently greater than λ_3, the optimal segmentation might give a *rent* (i.e. an additional profit) to the monopolist. By packing more consumers with lower types together, the social planner also makes higher types more distinguishable, thus allowing the monopolist to raise their prices. The above example illustrates the main argument of the paper: while market segmentation can redistribute surplus without any loss of efficiency, sometimes raising the surplus of poorer consumers can only be done if some of the surplus from richer consumers is granted to the monopolist.

However, not every aggregate market requires the granting of rents to the monopolist in order to satisfy redistributive objectives. Consider for instance the aggregate market $\mu^0 = (0.2, 0.65, 0.15)$, represented in the left panel of figure 1.5. The optimal segmentation of this market given *any* preferences $\lambda_2 \geq \lambda_3$ is the one depicted in the

figure: it always generates a segment with $\{v_1, v_2\}$ and another one with $\{v_2, v_3\}$, and this segmentation is always uniformly weighted consumer-optimal. On this aggregate market, satisfying a redistributive objective never requires granting rents to the monopolist because it contains sufficiently many consumers of type v_2, such that even after pooling as many as possible of them with types v_1 in segment μ, there are still sufficiently many types v_2 left to ensure that types v_3 will not be over-represented in segment μ'.

The result below characterizes the set of aggregate markets that, under a sufficiently strong redistributive motive, would require granting rents to the monopolist. We denote this set as the *rent region.*

Result 2. *The rent region is give by*

$$\mathrm{int}\Big(\mathrm{co}\Big(\{\delta_3, \mu^{123}, \mu^{12}, \mu^{23}\}\Big)\Big).$$

This result is illustrated in the right panel of figure 1.5, where the rent region is depicted in orange. Equivalently, the complement of this set denotes the aggregate markets for which any redistributive objective can be met without granting rents to the monopolist — that is, while maximizing total consumer surplus—. We call this set the *no-rent region.* The following section generalizes the insights presented

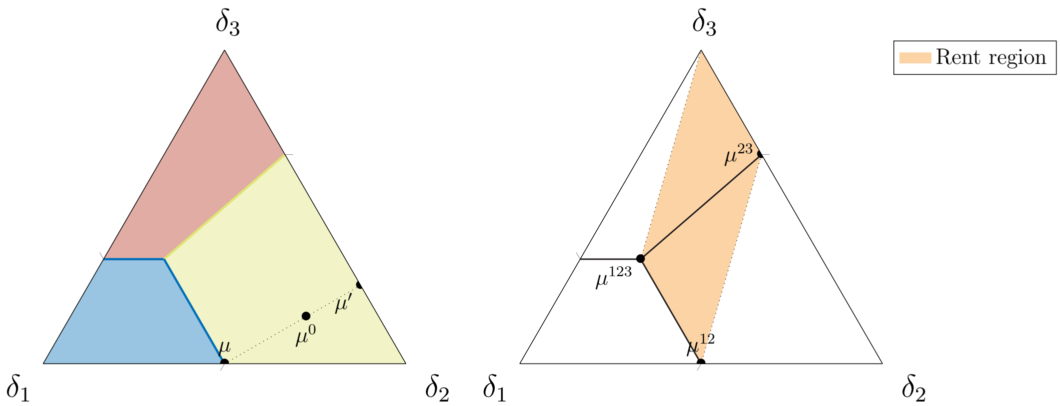


Figure 1.5: Rent Region.

through this example. Section 1.4.1 generalizes the fact that optimal segmentations are efficient and include discount segments supported at markets at which the monopolist is indifferent between more than one price, while section 1.4.2 establishes properties of optimal segmentations when the redistributive motive is sufficiently strong and shows how to construct optimal segmentations in this case. Finally, section 1.4.3 characterizes generally the no-rent and rent regions and shows that optimal segmentations for markets belonging to the no-rent region exhibit a very

simple form, with only one discount segment and one uniform price segment.

1.4 Optimal Segmentations

We now turn to the analysis of the general case. In section 1.4.1 we derive general properties of optimal segmentations — that is, characteristics that are present in optimal segmentations given any decreasing welfare weights λ. Section 1.4.2 then constructs optimal segmentations under strongly redistributive preferences: when the weight assigned to lower types is sufficiently larger than the weight assigned to higher types. Finally, we characterizes the set of aggregate markets for which satisfying a redistributive objective might require granting additional profits to the monopolist in section 1.4.3.

1.4.1 General Properties

Efficient segmentations. Our first result echoes our analysis of efficiency in the three-value case and establishes that i) we can always restrict ourselves to efficient segmentations—as long as the weights are non-negative; ii) if the weights are all strictly positive (i.e. if $\lambda_K > 0$ under our assumption of decreasing weights), only efficient segmentations can be optimal.

Proposition 1. *For any aggregate market μ^0 and any weights $\lambda \in \mathbb{R}^K_+$ (not necessarily decreasing), there exists an efficient optimal segmentation of μ^0. Furthermore, if every weight is strictly positive, then any optimal segmentation is efficient.*

Proof. This result is a direct consequence of Proposition 1 in Haghpanah and Siegel (2022b)—which itself follows from the proof of Theorem 1 in Bergemann et al. (2015). □

This result relies on the fact that any inefficient market can be segmented in a Pareto improving manner, that is, in a way that weakly increases the surplus of all consumers. Hence, as long as the social planner does not assign a negative weight to any consumer, there must be an efficient optimal segmentation. Proposition 1 thus implies that segmenting in a redistributive manner never comes at the expense of efficiency.

Direct segmentations. A segmentation σ is *direct* if all segments in σ have different prices, that is, if for any $\mu, \mu' \in \text{supp}(\sigma)$, $p(\mu) \neq p(\mu')$. Our next lemma shows that it is without loss of generality to focus on direct segmentations.

Lemma 1. *For any aggregate market μ^0 and any segmentation $\sigma \in \Sigma(\mu^0)$, there exists a direct segmentation $\sigma' \in \Sigma(\mu^0)$ such that,*

$$\int_{\Delta(M)} W(\mu)\,\sigma(\mathrm{d}\mu) = \int_{\Delta(M)} W(\mu)\,\sigma'(\mathrm{d}\mu).$$

Proof. See Appendix. □

We further show that there always exists an optimal and direct segmentation that is only supported on the boundaries of price regions $\{M_k\}_{k\in\mathcal{K}}$. Let $\mathcal{K}^0 := \{k \in \mathcal{K} \mid v_k \in \mathrm{supp}(\mu^0)\}$ be the set of indices of consumers' types supported by μ^0.

Lemma 2. *For any aggregate market μ^0 that is not efficient, there exists an optimal direct segmentation supported on boundaries of sets $\{M_k\}_{k\in\mathcal{K}^0}$.*

Proof. See Appendix. □

This result implies that we can restrict without loss of generality to finitely supported segmentations.

1.4.2 Strongly Redistributive Social Preferences

In this section, we derive some characteristics of the optimal segmentation when the social planner's preferences are *strongly redistributive*, that is, when the weights λ are strongly decreasing on the type v.

Definition 1. *The weights λ are κ-*strongly redistributive *if, for any $k < k' \le K-1$, $\frac{\lambda_k}{\lambda_{k'}} \ge \kappa$.*

That is, a social planner exhibits κ-strongly redistributive preferences (κ-SRP) if the weight she assigns to a consumer of type v_k is at least κ times larger than the weight she assigns to any consumer of type greater than v_k.

Let us define the *dominance* ordering between any two sets.

Definition 2. *Let $X, Y \subset \mathbb{R}$. The set X* dominates *Y, denoted $X \geqslant_D Y$, if for any $x \in X$ and any $y \in Y$, $x \ge y$.*[6]

We can now state the main result of this section.

Proposition 2. *For any aggregate market μ^0 in the interior of M, there exists $\underline{\kappa}$ such that if λ's are $\underline{\kappa}$-strongly redistributive, then for any optimal direct segmentation $\sigma \in \Sigma(\mu^0)$ and any markets $\mu, \mu' \in \mathrm{supp}(\sigma), \mu \neq \mu'$: either $\mathrm{supp}(\mu) \geqslant_D \mathrm{supp}(\mu')$ or $\mathrm{supp}(\mu') \geqslant_D \mathrm{supp}(\mu)$.*

[6] Note that this definition of dominance is stronger than the strong set order in Topkis (1998).

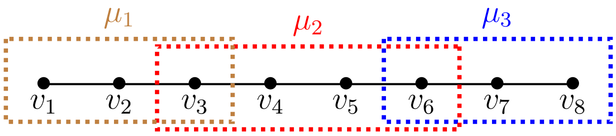


Figure 1.6: Structure of optimal segmentations under strong redistributive preferences.

Proof. See Appendix. □

The result stated above establishes that, when the social planner's preferences exhibit a sufficiently strong taste for redistribution, optimal segmentations divide the type space V into contiguous overlapping intervals, with the overlap between any two segments being composed of at most one type. The following corollary is a direct consequence of proposition 2:

Corollary 1. *For any aggregate market μ^0 in the interior of M, there exists $\underline{\kappa}$ such that if λ's are $\underline{\kappa}$-strongly redistributive, then for any optimal direct segmentation $\sigma \in \Sigma(\mu^0)$, any market $\mu \in \text{supp}(\sigma)$ and any k such that $\min\{\text{supp}(\mu)\} < v_k < \max\{\text{supp}(\mu)\}$: $\sigma(\mu)\mu_k = \mu^0_k$.*

The above result states that any segment μ belonging to a segmentation that is optimal under strong redistributive preferences contains *all* of the consumers with types strictly in-between $\min\{\text{supp}(\mu)\}$ and $\max\{\text{supp}(\mu)\}$. Together with proposition 2, it implies that, under κ-SRP optimal segmentations, every consumer type v will belong to *at most* two segments: either it will belong to the interior of the support of a segment μ, such that all consumers of this type have surplus $v - \min(\text{supp}(\mu))$, or it will be the boundary type between two segments μ and μ', such that a fraction of these consumers (those belonging to segment μ) gets surplus $v - \min(\text{supp}(\mu))$ and the rest gets no surplus. The structure of optimal segmentations under strong redistributive preferences is illustrated in figure 1.6.

These results, along with proposition 1, completely pin down the κ-SRP optimal direct segmentation. One can construct it by employing the following procedure, presented as follows through steps:

- *Step i)* Start by creating a segment — call it μ_a — with all consumers of type v_1.
- *Step ii)* Proceed to including in μ_a, successively, all consumers of type v_2, then all of the types v_3, and so on. From proposition 1 we know that μ_a must be efficient, meaning that we must have $p(\mu_a) = v_1$. As such, the process of inclusion of types higher than v_1 must be halted at the point in which adding

a new consumer in μ_a would result in v_1 no longer being an optimal price in this segment. We denote as $v_{(a|b)}$ the type that was being included when the process was halted.

- *Step iii)* Create a new segment — call it μ_b — with all of the remaining types $v_{(a|b)}$.
- *Step iv)* Proceed to including in μ_b, successively, all of consumers of type $v_{(a|b)+1}$, then all of the types $v_{(a|b)+2}$, and so on. Halt this process at the point in which adding a new consumer in μ_b would result in $v_{(a|b)}$ no longer being an optimal price in this segment. We denote as $v_{(b|c)}$ the type that was being included when the process was halted.
- *Step v)* Create a new segment with all of the remaining types $v_{(b|c)}$. Repeat the process described in the last steps until every consumer has been allocated to a segment.

1.4.3 Optimal Segmentations and Informational Rents

This section explores the question of when does an optimal segmentation maximize total consumer surplus or, conversely, when it grants a rent for the monopolist.

Say that an aggregate market μ^0 belongs to the *rent region* if there exists some $\underline{\kappa}$ such that if the social planner has $\underline{\kappa}$-strongly redistributive preferences, the optimal segmentation grants a rent to the monopolist. Conversely, denote *no-rent region* the set of aggregate markets for which any optimal segmentation with redistributive preferences also maximizes total consumer surplus.

Before we characterize the rent and no-rent regions, we define a particular segmentation, which we will call σ^{NR}:

Definition 3. *Let μ^0 be an aggregate market with uniform price v_u. Call σ^{NR} the segmentation that splits μ^0 into two segments μ^s and μ^r, such that:*

$$\mu^s = \left(\frac{\mu_1^0}{\sigma}, \frac{\mu_2^0}{\sigma}, \ldots, \mu_u^s, 0, \ldots, 0\right),$$
$$\mu^r = \left(0, 0, \ldots, \mu_u^r, \frac{\mu_{u+1}^0}{1-\sigma}, \ldots, \frac{\mu_K^0}{1-\sigma}\right),$$

where $\mu_u^s = v_1/v_u$, $\mu_u^r = (\mu_u^0 - \sigma\mu_u^s)/(1-\sigma)$ and $\sigma = (v_u \sum_{i=1}^{u-1} \mu_i^\star)/(v_u - v_1)$.

Segmentation σ^{NR} is very simple and generates only two segments: one pooling all the consumers who would not buy the good on the unsegmented market (those

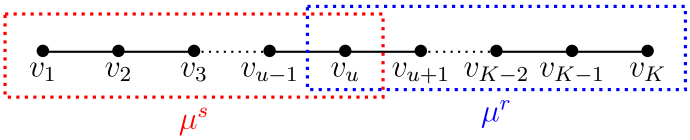


Figure 1.7: Segmentation σ^{NR}.

with type lower than v_u) and another one pooling all the consumers who would buy the good on the unsegmented market (those with type higher than v_u). Under segmentation σ^{NR}, the only consumer type that gets assigned to two different segments is v_u.

Proposition 3. *An aggregate market μ^0 belongs to the no-rent region if and only if σ^{NR} is an efficient segmentation of μ^0.*

Proof. See Appendix. □

Proposition 3 establishes a simple criterion that defines whether an aggregate market belongs to the no-rent region: it suffices to check if, under σ^{NR}, $p(\mu^s) = v_1$ and $p(\mu^r) = v_u$. Whenever this is not true, the aggregate market belongs to the rent region.

Corollary 2. *Consider an aggregate market μ^0. If σ^{NR} is not an efficient segmentation of μ^0, then there exists $\underline{\kappa}$ such that, if welfare weights λ are $\underline{\kappa}$-strongly redistributive, any optimal segmentation grants a rent to the monopolist.*

The intuition for the results above is as follows. A market belongs to the no-rent region if, given any redistributive preferences, its optimal segmentation maximizes total consumer surplus. On one hand, we know from proposition 2 that, under strong redistributive preferences, optimal segmentations divide the type space into overlapping intervals, with the overlap between two segments being comprised of at most one type. On the other hand, we have as a necessary and sufficient condition for total consumer surplus to be maximized that the segmentation is i) efficient and ii) the uniform price v_u is an optimal price at *every* segment generated by this segmentation. Condition i) ensures that total surplus is maximized, while condition ii) ensures that producer surplus is kept at it's uniform price level, meaning that all of the surplus created by the segmentation goes to consumers. Since condition ii) can only be satisfied if type v_u belongs in the support of all segments, we get that the conditions for optimality under strong redistributive preferences and for total consumer surplus to be maximized can only be simultaneously met by a segmentation that only generates two segments, with the overlap in the support of both segments being comprised of v_u.

Such a segmentation indeed maximizes total consumer surplus if it is efficient and if v_u is an optimal price in both segments. This is the case if v_1 and v_u are both optimal optimal prices on the lower segment, and if v_u is an optimal price in the upper segment. Segmentation σ^{NR} is the *only* segmentation that can potentially satisfy all of these conditions at once, as it includes in the lower segment the exact proportion of types v_u that would make the monopolist indifferent between charging a price of v_1 or v_u. As such, segmentation σ^{NR} maximizes total consumer surplus if and only if it is efficient.

Corollary 3. *If an aggregate market μ^0 belongs to the no-rent region, then σ^{NR} is its only direct consumer-optimal segmentation under any redistributive preferences.*

This result establishes that, for markets in the no-rent region, optimal segmentations have an extremely simple structure: they only generate a discount segment with price v_1, pooling all the types who would not consume under the uniform price and some of the types v_u, and a residual segment with price v_u, containing all of the remaining consumers. Furthermore, this segmentation must be optimal under *any* decreasing welfare weights λ. As such, this result selects for the markets belonging to the no-rent region one among the many uniformly weighted consumer-optimal segmentations that were outlined in Bergemann et al. (2015).

Due to the structure of segmentation σ^{NR}, all of the surplus that is generated by the segmentation is given to consumers with types below or equal to v_u, all of which get the maximum surplus they could potentially get. Since it is impossible to raise the surplus of any type below v_u, and impossible to raise the surplus of types above v_u without redistributing from lower to higher types, this segmentation must be optimal whenever the weights assigned to different consumers are (weakly) decreasing on the type.

The results in this section establish that there are essentially two types of markets: those for which redistribution can be done only within consumers, while keeping total consumer surplus maximal, and those for which increasing the surplus of lower types past a certain point necessarily decreases the total pie of surplus accruing to consumers and grants additional profits to the monopolist.

Bibliography

Akbarpour, M., Dworczak, P., and Kominers, S. D. (2020). Redistributive allocation mechanisms. *Working Paper.*

Ali, S. N., Lewis, G., and Vasserman, S. (2022). Voluntary Disclosure and Personalized Pricing. *The Review of Economic Studies.* rdac033.

Bergemann, D., Brooks, B., and Morris, S. (2015). The limits of price discrimination. *American Economic Review*, 105(3):921–57.

Dube, J.-P. and Misra, S. (2022). Personalized pricing and consumer welfare. *Journal of Political Economy*, Forthcoming.

Dworczak, P., Kominers, S. D., and Akbarpour, M. (2021). Redistribution Through Markets. *Econometrica*, 89(4):1665–1698.

Elliott, M., Galeotti, A., Koh, A., and Li, W. (2021). Market segmentation through information. *Working Paper.*

Haghpanah, N. and Siegel, R. (2022a). The limits of multi-product price discrimination. *American Economic Review: Insights*, Forthcoming.

Haghpanah, N. and Siegel, R. (2022b). Pareto improving segmentation of multi-product markets. *Journal of Political Economy*, Forthcoming.

Hidir, S. and Vellodi, N. (2020). Privacy, Personalization, and Price Discrimination. *Journal of the European Economic Association*, 19(2):1342–1363.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory.* Oxford University Press, New York.

Pai, M. and Strack, P. (2022). Taxing Externalities Without Hurting the Poor. *Working Paper.*

Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.

Ravid, D., Roesler, A.-K., and Szentes, B. (2022). Learning before trading: On the inefficiency of ignoring free information. *Journal of Political Economy*, 130(2):346–387.

Robinson, J. (1933). *The Economics of Imperfect Competition*. London: Macmillan.

Roesler, A.-K. and Szentes, B. (2017). Buyer-optimal learning and monopoly pricing. *American Economic Review*, 107(7):2072–80.

Schmalensee, R. (1981). Output and welfare implications of monopolistic third-degree price discrimination. *American Economic Review*, 71(1):242–247.

Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton University Press.

Varian, H. R. (1985). Price discrimination and social welfare. *American Economic Review*, 75(4):870–875.

Chapter 2

Persuading a Wishful Thinker

Abstract

We analyze a model of persuasion in which Receiver forms wishful non-Bayesian beliefs. The effectiveness of persuasion depends on Receiver's material stakes: it is more effective when intended to encourage risky behavior that potentially leads to a high payoff and less effective when intended to encourage more cautious behavior. We illustrate this insight with applications showing why informational interventions are often ineffective in inducing greater investment in preventive health treatments, how financial advisors might take advantage of their clients overoptimistic beliefs and why strategic information disclosure to voters with different partisan preferences can lead to belief polarization in an electorate.

JEL classification codes: D82; D83; D91.
Keywords: non-Bayesian persuasion; motivated thinking; overoptimism; optimal beliefs.

---

[0]This chapter is joint work with Victor Augias and previously circulated under the title "Wishful Thinking: Persuasion and Polarization." This version of the paper is the one submitted to the journal Games & Economic Behavior on February 28, 2022, currently at the "Reject and Resubmit" stage. We would like to point out that the present version of this chapter is being extensively modified in view of the resubmission of the paper. We thank Jeanne Hagenbach and Eduardo Perez-Richet for their support. We also thank S. Nageeb Ali, Roland Bénabou, Michele Fioretti, Alexis Ghersengorin, Simon Gleyze, Emeric Henry, Deniz Kattwinkel, Frédéric Koessler, Laurent Mathevet, Meg Meyer, Daniel Monte, Nikhil Vellodi, Adrien Vigier and Yves Le Yaouanq for their valuable feedbacks and comments, as well as seminar audiences at Sciences Po, Paris School of Economics, São Paulo School of Economics (FGV) and at the Econometric Society European Meeting 2021. All remaining errors are ours. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 850996 – MOREV and 101001694 – IMEDMC).

2.1 Introduction

It is generally assumed in models of strategic communication that receivers update beliefs in a perfectly rational manner, as would a Bayesian statistician. Yet, a substantial literature in psychology and behavioral economics shows that the process by which individuals interpret information and form beliefs is not guided solely by a desire for accuracy but often depends on their motivations and material incentives. This phenomenon is generally referred to as *motivated inference* (Kunda, 1987, 1990), and a common manifestation of it is *wishful thinking*: the tendency of individuals to let their *preferences about outcomes* influence the way they process information, leading to beliefs that are systematically biased towards outcomes they wish to be true.[1] In this paper we investigate how wishful thinking affects the effectiveness of persuasion, i.e., the probability or frequency with which a sender is able to induce a receiver to take her preferred action.

Following Caplin and Leahy (2019), we propose a model in which the receiver's belief updating rule is non-bayesian: after observing an informative signal, Receiver forms beliefs by trading off their anticipatory value against the psychological cost of distorting beliefs away from Bayesian ones. As a result, Receiver's beliefs are stakes-dependent, i.e., they depend on his preferences, and overweight the state associated with the highest payoff, giving rise to overoptimism.

Distortions in beliefs lead to distortions in Receiver's behavior: some actions end up being favored, meaning that they are taken more often (i.e., after the reception of a strictly greater set of possible signals) relative to a Bayesian decision-maker. When he only has two available actions, wishful thinking leads Receiver to favor the action associated with the highest payoff and the highest payoff variability. If one of the two actions induces the highest possible payoff and the other induces the highest payoff variability, then which of the two is favored depends on the magnitude of Receiver's belief distortion cost. As such, the effectiveness of information provision as a tool to incentivize agents might vary with individuals' material stakes: *persuasion is more effective when it is aimed at encouraging behavior that is risky but can potentially yield very high returns and less effective when it is aimed at encouraging more cautious behavior*. We illustrate this insight in applications in which wishful beliefs can play an important role.

[1] There exists abundant experimental evidence of wishful thinking. See in particular Bénabou and Tirole (2016), page 150 and Benjamin (2019) Section 9, as well as, e.g., Weinstein (1980), Mijović-Prelec and Prelec (2010), Mayraz (2011), Heger and Papageorge (2018), Coutts (2019), Engelmann et al. (2019) or Jiao (2020).

Application 1: Information Provision and Preventive Health Care. In this application a public health agency designs an information policy about the risk of infection of an illness in order to promote a preventive treatment that can be adopted by individuals at some cost. Since not adopting the treatment is the action that can potentially yield the highest payoff (in case the illness is not severe) and also the action with the highest payoff variability, it is favored by wishful receivers. As such, information campaigns aimed at promoting preventive behavior are less effective. We also show how the effectiveness of information campaigns are impacted by the severity of the disease and the effectiveness of the treatment.

This application sheds light on the stylized fact that individuals are consistently investing too little in preventive health care treatments, even if offered at low prices (especially in developing countries, see Dupas, 2011; Chandra et al., 2019; Kremer et al., 2019, Section 3.1) and that informational interventions are often ineffective in inducing more investment in preventive health care devices (see, in particular, Dupas, 2011, Section 4, and Kremer et al., 2019, Section 3.3). Recent literature conjectures that individuals might not be responsive to such information campaigns because they prefer to hold optimistic prospects about their health risks (see Schwardmann, 2019 and Kremer et al., 2019, Section 3.3).[2] Our model formalizes this argument.

Application 2: Persuading a Wishful Investor. In this application, we consider the interaction between a financial broker and her potential client. The broker designs reports about the (continuously distributed) return of some risky financial product to persuade the client to buy the asset. We show that a financial broker interested in selling a risky product is always more effective when persuading a wishful investor.

This application formalizes why some professional financial advisors might sometimes not act in the best interest of their clients by making investment recommendations that take advantage of their biases and mistaken beliefs (see, for instance, Mullainathan et al., 2012 or Beshears et al., 2018, Section 9) as well as why some consulting firms seem to specialize in advice misconduct and cater to biased consumers (Egan et al., 2019). It also helps explaining why the online betting industry puts so much effort into persuasion. Indeed, Babad and Katz (1991) document that individuals generally display wishful thinking when they take part in lotteries: they prefer to think they will win and are therefore more receptive to information encouraging risky bets.

[2] There exists compelling experimental evidence that such self-deception exists in the medical testing context (Lerman et al., 1998; Oster et al., 2013; Ganguly and Tasoff, 2017).

Application 3: Public Persuasion and Political Polarization. Belief polarization along partisan lines is a pervasive and much debated feature of contemporary societies. Although such polarization can be partly caused by differential access to information, evidence suggests that it is exacerbated by the fact that individuals tend to make motivated inferences about the *same* piece of information (Babad, 1995; Thaler, 2020).

In this application we explore the relationship between optimal information disclosure to wishful citizens and belief polarization. Following Alonso and Câmara (2016), we model a majority voting setting in which an electorate, differentiated in terms of partisan preferences, uses information disclosed by a politician to vote on a proposal. Wishful thinking leads voters with different preferences to adopt different beliefs after being exposed to a public signal: those voting against or for the proposal distort their beliefs in opposite directions, giving rise to polarization. Sender's optimal public experiment consists in persuading the median voter, which maximizes the number of voters distorting beliefs in opposite directions. We show that if partisan preferences are symmetrically distributed around the median, then Sender's optimal information policy generates maximal belief polarization in the electorate as a byproduct. This adds nuance to the argument that motivated thinking is one of the drivers of polarization: not only can motivated thinking lead to polarization, but the strategic disclosure of information to a motivated electorate can also accentuate this tendency[3].

2.1.1 Related literature

The persuasion and information design literature[4] has initially focused on the problem of influencing rational Bayesian decision-makers as in the seminal contributions of Kamenica and Gentzkow (2011) and Bergemann and Morris (2016). By introducing non-Bayesian updating in the form of motivated beliefs formation, we contribute to the literature studying persuasion of receivers subject to mistakes in probabilistic

[3]This application is related to the paper by Le Yaouanq (2021) who constructs a model of large elections with motivated voters. As in our model, the formation of motivated beliefs by citizens leads voters with different preferences to hold different beliefs after observing the same information. We find, as he does, that greater heterogeneity in partisan preferences increases belief polarization but has no effect on the policy implemented in equilibrium. This is, however, the consequence of a different modelling assumption. Namely, that information is endogenously designed to persuade the median voter, whose vote is not distorted relative to a Bayesian voter.

[4]See Bergemann and Morris (2019) and Kamenica (2019) for reviews of this literature.

inferences.[5][6] Levy et al. (2018) analyze a Bayesian persuasion problem where a sender can send multiple signals to a receiver subject to correlation neglect. Benjamin et al. (2019) provide an example of persuasion game where Receiver exhibits base-rate neglect when updating beliefs. In de Clippel and Zhang (2020) the receiver holds subjective beliefs which belong to a broader class of distorted Bayesian posteriors. In contrast, in our model, Receiver's belief formation process optimally trades-off the benefits and costs associated with maintaining non-Bayesian beliefs as in the work of Caplin and Leahy (2019).

On the one hand, we assume that Receiver's value from maintaining inaccurate beliefs comes from the anticipation of the payoff he will achieve in equilibrium. Intuitively, it represents the idea that individuals might derive utility from the anticipation of future outcomes, be them good or bad. This hypothesis has been widely used in the literature to study how anticipatory emotions affect physical choices (see, e.g., Loewenstein, 1987; Caplin and Leahy, 2001) as well as choices of beliefs (Bénabou and Tirole, 2002; Brunnermeier and Parker, 2005; Bracha and Brown, 2012; Caplin and Leahy, 2019). Receiver's choice of beliefs is thus a way of satisfying his psychological need to be optimistic about the best-case outcomes or, on the contrary, to avoid the dread and anxiety associated with the worst-case outcomes. This hypothesis is supported experimentally by Engelmann et al. (2019), who find significant evidence that wishful thinking is caused by the desire to reduce anxiety associated with anticipating bad events. It is important to note that while anticipatory utility may be a strong motive for manipulating one's beliefs, it is not the only possible one. This differentiates wishful thinking from the more general concept of motivated reasoning, which is usually defined as the degree to which individuals' cognition is affected by their motivations.[7] Different motivations from anticipated payoffs have been explored in the literature such as cognitive dissonance avoidance (Akerlof and Dickens, 1982; Golman et al., 2016), preference to believe in a "Just World" (Bénabou and Tirole, 2006), maintaining high motivation when individuals are aware of being subject to a form of time-inconsistency (Bénabou and

[5]See Benjamin (2019) for a review of the literature. In particular, wishful thinking belongs to preference-biased inferences reviewed in Benjamin (2019), Section 9.

[6]It is interesting to note that an active literature also explores how errors in strategic reasoning (Eyster, 2019) affect equilibrium outcomes in strategic communication games. Although in our model Receiver understands all the strategic issues, we believe, nevertheless, that it is important to mention that players' misunderstanding of their strategic environment might also lead them to make errors in statistical inference even if they update beliefs via Bayes' rule, as in Mullainathan et al. (2008), Ettinger and Jehiel (2010), Hagenbach and Koessler (2020) and Eliaz et al. (2021b,a) who consider communication games where players make inferential errors because of a coarse understanding of their environment.

[7]See Krizan and Windschitl (2009) for a more detailed discussion on the differences between wishful thinking and motivated reasoning.

Tirole, 2002, 2004) or satisfying the need to belong to a particular identity (Bénabou and Tirole, 2011).

On the other hand, we assume distorting beliefs away from the Bayesian benchmark is subject to some psychological cost. This assumption reflects the idea that, under a motivated cognition process (Kunda, 1987, 1990), individuals may use sophisticated mental strategies such as manipulating their own memory (Bénabou, 2015; Bénabou and Tirole, 2016)[8], avoiding freely available information (Golman et al., 2017) or creating elaborate narratives supporting their bad choices or inaccurate claims to justify their preferred beliefs.[9] Our assumptions on the cost function captures, in "reduced form", the fact that implementing such mental strategies comes at a cost when desired beliefs deviate from from the Bayesian rational ones. In contrast, Brunnermeier and Parker (2005) model the cost of erroneous beliefs as the instrumental loss associated with the inaccurate choices induced by such beliefs. It is worth noting that Coutts (2019) provides experimental evidence in favor of the psychological rather than instrumental costs associated with belief distortion.

2.2 Model

States and prior belief. A state of the world θ is drawn by Nature from a state space Θ according to a prior distribution $\mu_0 \in \mathrm{int}(\Delta(\Theta))$.[10] Receiver (he) and Sender (she) do not observe the state ex-ante but its prior distribution is common knowledge.

Actions and payoffs. Receiver chooses an action a from a compact space A with at least two actions. His material payoff is given by $u(a, \theta)$.[11] Receiver's choice affects Sender's payoff, which is given by $v(a)$. Before Receiver takes his action, Sender can commit to any signal structure (σ, S) given by an endogenously chosen set of signal realizations S and a stochastic mapping $\sigma \colon \Theta \to \Delta(S)$ associating any realized state θ to a conditional distribution $\sigma(\theta)$ over S.

[8]For experimental evidence on memory manipulation see, e.g., Saucet and Villeval (2019), Carlson et al. (2020) and Chew et al. (2020).

[9]One can relate this possible microfoundation of the belief distortion cost to the literature on lying costs (Abeler et al., 2014, 2019) since, when Receiver is distorting away his subjective belief from the rational Bayesian beliefs, he is essentially lying to himself. We thank Emeric Henry for suggesting us this interpretation of the cost function.

[10]In what follows, for any nonempty Polish space X, we denote $\Delta(X)$ the set of Borel probability measures over the measure space $(X, \mathcal{B}(X))$. We always endow $\Delta(X)$ with the weak*-topology. If the support of a measure $\mu \in \Delta(X)$ is finite we adopt the shorthand notation $\mu(\{x\}) = \mu(x)$ for any $x \in \mathrm{supp}(\mu)$.

[11]We assume the map $u(a, \cdot) \colon \Theta \to \mathbb{R}$ to be Borel measurable, continuous and bounded for any $a \in A$.

Receiver's behavior. For any belief $\eta \in \Delta(\Theta)$, Receiver's optimal action correspondance is given by

$$A(\eta) = \arg\max_{a \in A} \int_{\Theta} u(a, \theta)\, \eta(\mathrm{d}\theta).$$

Without loss of generality, we assume that no action is dominated, i.e., for any action $a \in A$ there always exists some belief η such that $a \in A(\eta)$. When the set $A(\eta)$ has more than one element we break the tie in favor of Sender. That is, for any belief η, the action played by Receiver in equilibrium is given by a selection $a(\eta) \in A(\eta)$ which maximizes Sender's expected payoff.[12]

Receiver's beliefs. After observing any signal realization $s \in S$, a Bayesian decision-maker's belief is given by

$$\mu\left(\tilde{\Theta}|s\right) = \frac{\int_{\tilde{\Theta}} \sigma(s|\theta)\, \mu_0(\mathrm{d}\theta)}{\int_{\Theta} \sigma(s|\theta)\, \mu_0(\mathrm{d}\theta)},$$

for any Borel set $\tilde{\Theta} \subseteq \Theta$.

In contrast, we assume that, when forming beliefs, Receiver trades-off the psychological benefit against the psychological cost of holding possibly non-Bayesian beliefs. The psychological benefit of Receiver under a certain belief η is given by his *anticipated material payoff*

$$U(\eta) = \int_{\Theta} u(a(\eta), \theta)\, \eta(\mathrm{d}\theta).$$

However, holding belief η when the Bayesian belief generated by some signal is μ comes at a psychological cost $C(\eta, \mu)$ for Receiver. We assume that this cost is given by the Kullback-Leibler divergence between η and μ, formally defined by

$$C(\eta, \mu) = \int_{\Theta} \frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta) \ln\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta)\right) \mu(\mathrm{d}\theta),$$

for any $\eta, \mu \in \Delta(\Theta)$, where $\mathrm{d}\eta/\mathrm{d}\mu$ is the Radon-Nikodym derivative of η with respect to μ, defined whenever η is absolutely continuous with respect to μ. This assumption is made for tractability but does not qualitatively affect our main results.[13]

[12] There might be more than one such selection if there exists some $\eta \in \Delta(\Theta)$ at which Sender is indifferent between some actions in $A(\eta)$. In that case, we pick arbitrarily one of those.

[13] We show that our results on Receiver's equilibrium beliefs and behavior continue to hold when the psychological cost functions belongs to a more general class of statistical divergences in

Accordingly, we define Receiver's *psychological payoff* as

$$\Psi(\eta, \mu) = U(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

for any $\eta, \mu \in \Delta(\Theta)$, where $\rho \in \mathbb{R}_+^*$ parametrizes the extent of Receiver's *wishfulness*. Receiver's belief η must maximize his psychological payoff given any Bayesian belief μ. Therefore, it must belong to the optimal beliefs correspondence

$$B(\mu) = \underset{\eta \in \Delta(\Theta)}{\arg\max}\ \Psi(\eta, \mu),$$

for any $\mu \in \Delta(\Theta)$, and Receiver's psychological payoff when he holds a belief $\eta \in B(\mu)$ is

$$\Psi(\mu) = \max_{\eta \in \Delta(\Theta)} \Psi(\eta, \mu),$$

for any Bayesian posterior $\mu \in \Delta(\Theta)$.[14] We assume that when Receiver is psychologically indifferent between several beliefs in $B(\mu)$ he picks the one that maximizes Sender's expected utility. Therefore, Receiver's *equilibrium belief* is given by a selection $\eta(\mu) \in B(\mu)$ which maximizes Sender's expected payoff.[15] This tie breaking rule ensures that the Receiver's equilibrium belief is uniquely defined and simplifies the characterization of the optimal information policy.

Persuasion problem. We can equivalently think of Sender committing ex-ante to a signal structure (σ, S) or to an *information policy* $\tau \in \mathcal{T}(\mu_0)$, where

$$\mathcal{T}(\mu_0) = \left\{ \tau \in \Delta\left(\Delta(\Theta)\right) : \int_{\Delta(\Theta)} \mu(\tilde{\Theta})\, \tau(\mathrm{d}\mu) = \mu_0(\tilde{\Theta}) \text{ for any Borel set } \tilde{\Theta} \subseteq \Theta \right\},$$

is the set of Bayes-plausible distributions over posterior beliefs given the prior μ_0.

We assume *Sender knows Receiver is a wishful thinker*. Accordingly, she correctly

---

appendix B.1.

[14]As already noted by Bracha and Brown (2012) as well as Caplin and Leahy (2019), this optimization problem has a similar mathematical structure to the multiplier preferences developed in Hansen and Sargent (2008) and axiomatized in Strzalecki (2011). Precisely, the agent in Strzalecki (2011) solves

$$\max_{a \in A} \min_{\eta \in \Delta(\Theta)} \int_{\Theta} u(a, \theta)\, \eta(\mathrm{d}\theta) + \frac{1}{\rho} C(\eta, \mu), \tag{2.1}$$

for any given $\mu \in \Delta(\Theta)$. In that model, the parameter ρ measures the degree of confidence of the decision-maker in the belief μ or, in other words, the importance he attaches to belief misspecification. Conclusions on the belief distortion in that setting are naturally reversed with respect to our model: a receiver forming beliefs according to equation (2.1) would form overcautious beliefs. Studying how a rational Sender would persuade a Receiver concerned by robustness seems an interesting path for future research.

[15]Again, if Sender is indifferent between some beliefs we pick arbitrarily one of those.

anticipates the belief Receiver holds in equilibrium. Since Receiver's equilibrium belief characterizes how he would distort his belief away from any realized Bayesian posterior, Sender can choose the best information policy by backward induction, knowing: (i) which belief $\eta(\mu)$ Receiver holds in equilibrium after a posterior $\mu \in \mathrm{supp}(\tau)$ is realized and (ii) which action $a(\eta(\mu))$ Receiver chooses in equilibrium given the distorted belief $\eta(\mu)$. Sender's indirect payoff function is therefore given by

$$v(\mu) = v\left(a(\eta(\mu))\right)$$

for any $\mu \in \Delta(\Theta)$ and, hence, Sender's value from persuading a wishful Receiver under the prior μ_0 is

$$V(\mu_0) = \max_{\tau \in \mathcal{T}(\mu_0)} \int_{\Delta(\Theta)} v(\mu)\, \tau(\mathrm{d}\mu). \tag{2.2}$$

2.3 Receiver's wishful beliefs and behavior

In this section, we first extend Caplin and Leahy (2019) results by characterizing Receiver's equilibrium beliefs and behavior without imposing any restrictions on the action or state space.

To begin with, let Receiver's anticipated material payoff under action a and belief η be defined by

$$U_a(\eta) = \int_\Theta u(a, \theta)\, \eta(\mathrm{d}\theta).$$

Moreover, let

$$\eta_a(\mu) = \arg\max_{\eta \in \Delta(\Theta)} U_a(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

be Receiver's belief motivated by action a under posterior μ and

$$\Psi_a(\mu) = \max_{\eta \in \Delta(\Theta)} U_a(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

be Receiver's maximal psychological payoff motivated by action a under posterior μ. We identify Receiver's equilibrium belief $\eta(\mu)$ by: (i) finding the belief motivated by action a under μ, resulting in psychological payoff $\Psi_a(\mu)$, for any a and μ; (ii) finding which action it is optimal to motivate by maximizing $\Psi_a(\mu)$ with respect to a. proposition 4 characterizes $\eta_a(\mu)$ and $\Psi_a(\mu)$ in closed-form.

Proposition 4. *Receiver's maximal psychological payoff motivated by action a under*

the Bayesian posterior μ is given by

$$\Psi_a(\mu) = \frac{1}{\rho} \ln \left(\int_{\Theta} \exp\left(\rho u(a,\theta)\right)\, \mu(\mathrm{d}\theta) \right), \tag{2.3}$$

and is attained uniquely at the belief

$$\eta_a(\mu)(\tilde{\Theta}) = \frac{\displaystyle\int_{\tilde{\Theta}} \exp\left(\rho u(a,\theta)\right)\, \mu(\mathrm{d}\theta)}{\displaystyle\int_{\Theta} \exp\left(\rho u(a,\theta)\right)\, \mu(\mathrm{d}\theta)}. \tag{2.4}$$

for any Borel set $\tilde{\Theta} \subseteq \Theta$.

Proof. See appendix B.1. □

Remark now that if the action a uniquely maximizes Receiver's psychological payoff under Bayesian posterior μ we have $\eta(\mu) = \eta_a(\mu)$. If, on the other hand, $\Psi_a(\mu) = \Psi_{a'}(\mu)$ at μ for some $a' \neq a$, meaning that Receiver is psychologically indifferent between two beliefs, then Sender breaks the tie. As a consequence, if $\mu \in \Delta(\Theta)$ satisfies

$$\Psi_a(\mu) > \Psi_{a'}(\mu), \tag{2.5}$$

for all $a' \neq a$, meaning that Receiver psychologically prefers action a to any other action a', then Receiver's equilibrium belief is given by

$$\eta(\mu)(\tilde{\Theta}) = \eta_a(\mu)(\tilde{\Theta}),$$

for any Borel set $\tilde{\Theta} \subseteq \Theta$. If $\mu \in \Delta(\Theta)$ satisfies

$$\Psi_a(\mu) = \Psi_{a'}(\mu),$$

for some $a' \neq a$, meaning that Receiver is psychologically indifferent between some actions a' and a, then Sender picks her preferred belief given by

$$\eta(\mu)(\tilde{\Theta}) = \eta_{a^*}(\mu)(\tilde{\Theta}),$$

where $a^* \in \arg\max_{\tilde{a} \in \{a,a'\}} v(\tilde{a})$.

First, we can see from equation (2.4) that Receiver only distorts beliefs that induce actions with state-dependant payoffs, i.e., Receiver's beliefs are *stakes-dependent*. Formally, for any $a \in A$, we have $\eta_a(\mu) \neq \mu$ if, and only if, there exists $\theta \neq \theta'$ such that $u(a,\theta) \neq u(a,\theta')$. Second, Receiver forms beliefs that overweight the states associated with the highest payoff, giving rise to *overoptimism*. Formally, we always

have $\eta_a(\mu)(\Theta_a) \geq \mu(\Theta_a)$ for any $a \in A$ where $\Theta_a = \arg\max_{\theta\in\Theta} u(a,\theta)$. Moreover, Receiver's belief about payoff maximizing states $\eta_a(\mu)(\Theta_a)$ grows monotonically and eventually converges to 1 as Receiver's wishfulness ρ grows from 0 to $+\infty$.[16]

As proposition 4 shows, wishful thinking leads Receiver to hold overoptimistic beliefs. The next result shows that wishful thinking distorts Receiver's behavior accordingly.

Corollary 4. *Under his equilibrium belief, Receiver's optimal action correspondence is given by*

$$A(\eta(\mu)) = \arg\max_{a\in A} \int_{\Theta} \exp\left(\rho u(a,\theta)\right) \, \mu(\mathrm{d}\theta),$$

for any $\mu \in \Delta(\Theta)$ so Receiver's equilibrium action $a(\eta(\mu))$ corresponds to Sender's preferred selection in $A(\eta(\mu))$.

Remark that this result comes as a direct consequence of proposition 4 as, by definition, any action a is optimal under the belief motivated by action a. As already observed by Caplin and Leahy (2019), the previous result states, in essence, that a Receiver forming wishful beliefs behaves as a Bayesian agent whose preferences are distorted by the function $z \mapsto \exp(\rho z)$ for any $z \in \mathbb{R}$. Importantly, from Sender's point of view, a wishful Receiver's behavior is indistinguishable from that of a Bayesian rational agent with payoff function $\exp(\rho u(a,\theta))$. Accordingly, since the function $z \mapsto \exp(\rho z)$ is strictly convex as soon as $\rho > 0$, an agent forming wishful beliefs is less risk averse than his Bayesian self.

Corollary 4 also shows that wishful thinking materializes in the form of "motivated errors" in the sense of Exley and Kessler (2019): by choosing psychologically desirable beliefs, Receiver commits systematic errors in his decision-making, i.e., acts as if he had cognitive limitations or behavioral biases relatively to a Bayesian decision-maker.

2.4 Sender's value from persuasion

In this section, we assume that the action space of Receiver is binary, so $A = \{0,1\}$, and that Sender wants to induce $a = 1$, so $v(a) = a$. We provide necessary and sufficient conditions on Receiver's preferences under which he would take action 1 under a greater set of beliefs than a Bayesian Receiver. This allows us to compare Sender's value from persuading a wishful rather than a Bayesian Receiver as a function of the model's primitives, that is: Receiver's preferences and wishfulness.

[16] This property comes from the fact that wishful beliefs take the form of a soft-max function. For the sake of completeness we provide a proof of this result in appendix B.2.

The restriction to a binary set of actions is with loss of generality but allows better tractability.

We start by defining the two following sets of beliefs:

$$\Delta_a^B = \{\mu \in \Delta(\Theta) : a \in A(\mu)\},$$

and

$$\Delta_a^W = \{\mu \in \Delta(\Theta) : a \in A(\eta(\mu))\},$$

for any $a \in A$. The set Δ_a^B (resp. Δ_a^W) is the subset of posterior beliefs supporting an action a as optimal for a Bayesian (resp. wishful) Receiver. We say that an action is *favored* by a wishful receiver if that action is supported as optimal on a strictly larger set of posterior beliefs by a wishful Receiver compared to a Bayesian.

Definition 4 (Favored action)**.** *An action $a \in A$ is favored by a wishful Receiver if $\Delta_a^B \subset \Delta_a^W$.*

Assume for now on that $\Theta = \{\underline{\theta}, \overline{\theta}\}$. We first characterize when a wishful Receiver favors action $a = 1$ when the state space is binary and show afterwards that our results extend to any finite state space. Let us denote $u(a, \underline{\theta}) = \underline{u}_a$ and $u(a, \overline{\theta}) = \overline{u}_a$ for any $(a, \theta) \in A \times \Theta$. Assume that Receiver wants to "match the state," such that $\overline{u}_1, \underline{u}_0 > \overline{u}_0, \underline{u}_1$. Define the *payoff variability under action* 0 by $u_0 = \underline{u}_0 - \overline{u}_0$, the *payoff variability under action* 0 by $u_1 = \overline{u}_1 - \underline{u}_1$ and the indicator of the *highest achievable payoff* by $u_{\max} = \underline{u}_0 - \overline{u}_1$. With a small abuse of notation, denote $\eta = \eta(\overline{\theta})$ and $\mu = \mu(\overline{\theta})$.

By corollary 4, comparing how a wishful Receiver behaves compared to a Bayesian one is equivalent to comparing the behavior of two Bayesian receivers with respective payoff functions $\exp(\rho u(a, \theta))$ and $u(a, \theta)$. Thus, denote μ^B (resp. $\mu^W(\rho)$) the belief at which a Receiver with preferences $u(a, \theta)$ (resp. $\exp(\rho u(a, \theta))$) is indifferent between the two actions. Those beliefs are respectively equal to

$$\mu^B = \frac{\underline{u}_0 - \underline{u}_1}{\underline{u}_0 - \underline{u}_1 + \overline{u}_1 - \overline{u}_0}$$

and

$$\mu^W(\rho) = \frac{\exp(\rho \underline{u}_0) - \exp(\rho \underline{u}_1)}{\exp(\rho \underline{u}_0) - \exp(\rho \underline{u}_1) + \exp(\rho \overline{u}_1) - \exp(\rho \overline{u}_0)}.$$

With only two states, a wishful Receiver favors action $a = 1$ if and only if $\mu^W < \mu^B$, since whenever that condition is satisfied a wishful Receiver takes action $a = 1$ under a larger set of beliefs than a Bayesian. Next proposition characterizes when this is the case.

Lemma 3. *Action $a = 1$ is favored by a wishful Receiver if, and only if:*

(i) $u_{\max} \leq 0$ *and* $u_0 < u_1$*, or;*

(ii) $u_{\max} < 0$*,* $u_0 > u_1$ *and* $\rho > \overline{\rho}$*, or;*

(iii) $u_{\max} > 0$*,* $u_0 < u_1$ *and* $\rho < \overline{\rho}$*.*

where $\overline{\rho}$ *is a strictly positive threshold such that*

$$\mu^W(\overline{\rho}) = \mu^B.$$

Proof. See appendix B.3. □

Two key aspects of Receiver's material payoff thus determine which action he favors: *the highest achievable payoff* as well as *the payoff variability* for both actions. It is easy to grasp the importance of the highest payoff. Since the wishful thinker always distorts his beliefs in the direction of the most favorable outcome, in the limit, when there is no cost of distorting the Bayesian belief, Receiver would fully delude himself and always play the action that potentially yields such a payoff. The payoff variability u_a, on the other hand, is precisely Receiver's marginal psychological benefit from distorting his belief under action a. Hence, the higher the payoff variability associated with action a, the more the uncertainty about θ is relevant when such action is played and the bigger the marginal gain in anticipatory payoff the wishful thinker would get from distorting beliefs.

lemma 3 states that if an action a has both the highest payoff $\underline{u}_0$ or $\overline{u}_1$ and the greatest payoff variability u_a among all actions $a \in A$, it is always favored. If an action has either the highest payoff or the greatest payoff variability, then the wishfulness parameter ρ defines whether or not it is favored: for high wishfulness the action with the highest payoff is favored, whereas for low wishfulness it is the action with the greatest payoff variability that is favored. The intuition is the following: for sufficiently high values of Receiver's wishfulness, Receiver can afford stronger overoptimism about the most desired outcome, thus favoring the action that potentially yields this outcome despite such action not being associated with the highest marginal psychological benefit. In contrast, for sufficiently low values of ρ, Receiver cannot afford too much overoptimism about the most desired outcome. Hence, he prefers to distort beliefs at the margin that yields the highest marginal psychological benefit, such that the action associated with the highest payoff variability is favored.

The next proposition extends lemma 3 to an arbitrary finite number of states.

Proposition 5. *Assume Θ is a finite set with more than two elements. Receiver favors action $a = 1$ if, and only if, for any pair of states $\theta, \theta' \in \Theta$, Receiver's material payoffs associated with those states and his wishfulness parameter ρ satisfy one of the conditions (i), (ii) or (iii) in lemma 3.*

Proof. See appendix B.4. $\square$

Proposition 5 can easily be visualized graphically in an example with three states. Assume $\Theta = \{0, 1, 2\}$ and denote $\mu^B_{\theta,\theta'}$ (resp. $\mu^W_{\theta,\theta'}$) the belief making a Bayesian (resp. wishful) Receiver indifferent between actions $a = 0$ and $a = 1$ when $\mu(\theta), \mu(\theta') > 0$ but $\mu(\theta'') = 0$ for any $\theta, \theta', \theta'' \in \Theta$. In figure 2.1 we illustrate how Δ^W_1 compares to Δ^B_1 when Receiver's payoff function is given by:

$u(a,\theta)$	$\theta = 0$	$\theta = 1$	$\theta = 2$
$a = 0$	2	3	-1
$a = 1$	1	0	4

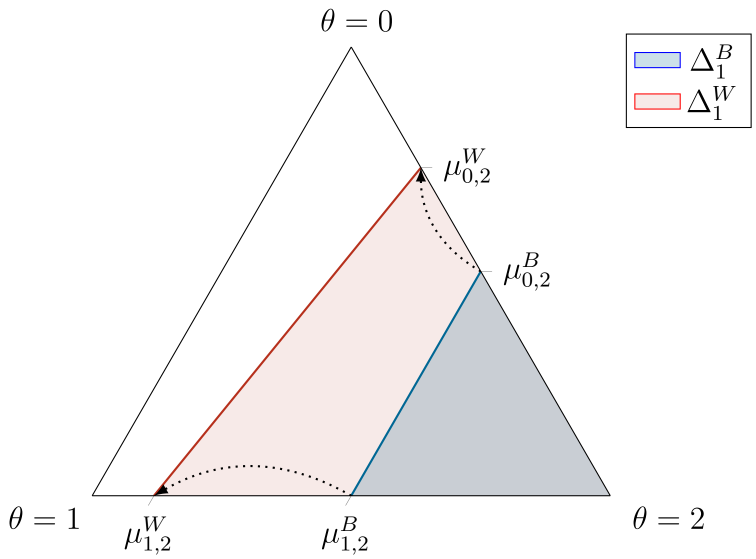


Figure 2.1: Comparison of supporting sets of beliefs. In blue, the set of Bayesian posteriors supporting action $a = 1$ for a Bayesian Receiver. In red, the set of Bayesian posteriors supporting action $a = 1$ for a wishful Receiver.

Notice that for the two pairs of states $(0, 2)$ and $(1, 2)$, the associated payoffs satisfy property (i) in lemma 3. That is, action $a = 1$ is associated with the highest payoff $u(1, 2) = 4$ as well as the highest payoff variability $u(1, 2) - u(0, 2) = 5$, under both pair of states. As a consequence, lemma 3 applies whenever focusing on those two pairs of states letting the other one being assigned probability zero. Then, we have $\mu^W_{0,2} > \mu^B_{0,2}$ and $\mu^W_{1,2} > \mu^B_{1,2}$. Remark now, that $\Delta^B_1 = \text{co}(\{\mu^B_{0,2}, \mu^B_{1,2}, \delta_2\})$ and

$\Delta_1^W = \text{co}(\{\mu^W_{0,2}, \mu^W_{1,2}, \delta_2\})$, where δ_θ denotes the Dirac distribution on state $\theta \in \Theta$. Consequently, $\Delta_1^B \subset \Delta_1^W$ so action $a = 1$ is favored by Receiver. If one of the conditions highlighted in lemma 3 were not satisfied for at least one of the pairs of states $(0, 2)$ or $(1, 2)$ then one of the thresholds $\mu^W_{\theta,\theta'}$ would be less or equal than $\mu^B_{\theta,\theta'}$ in which case Δ_1^W would not be a superset of Δ_1^B anymore.

Let us now turn our attention to the following questions: when is Sender better-off facing a wishful Receiver compared to a Bayesian and how does the (Blackwell) informativeness of Sender's optimal policy compare when persuading a wishful or a Bayesian Receiver? Remember that Sender chooses an information policy $\tau \in \Delta(\Delta(\Theta))$ maximizing

$$\int_{\Delta(\Theta)} v(\mu)\, \tau(\mathrm{d}\mu),$$

where

$$v(\mu) = \begin{cases} 1 & \text{if} \quad \mu \in \Delta_1^W \\ 0 & \quad \text{otherwise} \end{cases},$$

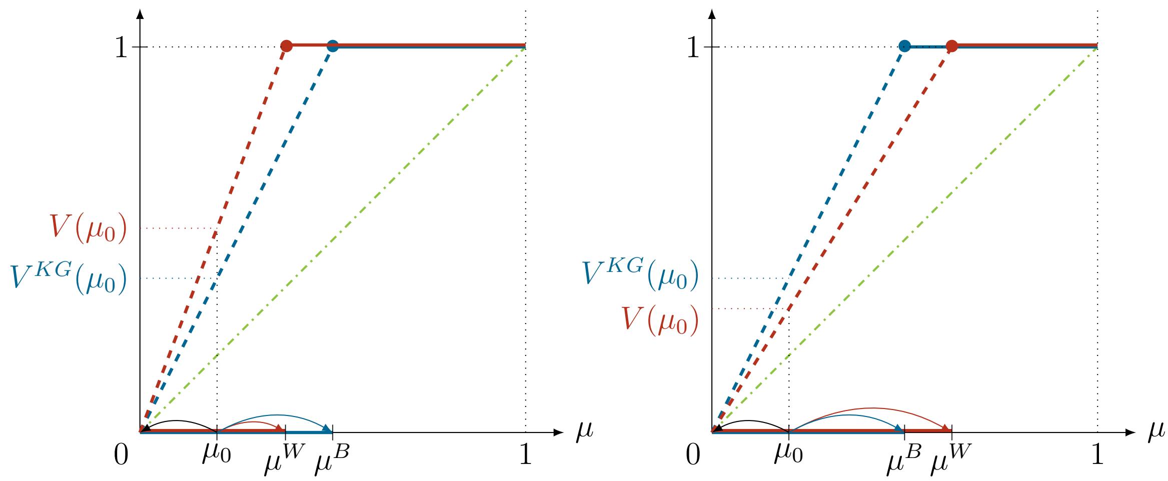
subject to the Bayes plausibility constraint

$$\int_{\Delta(\Theta)} \mu\, \tau(\mathrm{d}\mu) = \mu_0.$$

In the binary state case, it means that the threshold belief μ^W corresponds to the *smallest Bayesian posterior Sender needs to induce to persuade a wishful Receiver to take action* $a = 1$. Therefore, lemma 3 and proposition 5 have immediate consequences for Sender.

Corollary 5. *Let Θ be an arbitrary finite space with at least two elements. Then, Sender always achieves a weakly higher payoff when interacting with a wishful Receiver compared to a Bayesian for any prior $\mu_0 \in]0, 1[$ if, and only if, for any pair of states $\theta, \theta' \in \Theta$, Receiver's material payoffs associated with those states and his wishfulness parameter ρ satisfy one of the conditions (i), (ii) or (iii) in lemma 3. Moreover, when the state space is binary, Sender's optimal information policy is always weakly less (Blackwell) informative than in the Bayesian case.*

To illustrate corollary 5 we represent in figure 2.2 the concavifications of Sender's indirect utility when Receiver is wishful or Bayesian in two different cases. The case corresponding to lemma 3 is represented in figure 2.2a. Sender is always better-off persuading a wishful compared to a Bayesian receiver as $V(\mu_0) \geq V^{KG}(\mu_0)$ for any $\mu_0 \in]0, 1[$. On the other hand, if Receiver's preferences or wishfulness do not satisfy any of the properties in lemma 3, then Sender is weakly worse-off under any prior. This case is represented on figure 2.2b.



(a) At least one property in lemma 3 is satisfied.

(b) No property in lemma 3 is satisfied.

Figure 2.2: Expected payoffs under optimal information policies. Red curves: expected payoffs under wishful thinking. Blue curves: expected payoffs when Receiver is Bayesian. Dashed-dotted green lines: expected payoffs under a fully revealing experiment.

When Sender wants to induce an action that is (resp. is not) favored by a wishful Receiver, persuasion is always "easier" (resp. "harder") for Sender in the following sense: Sender needs a strictly less (resp. strictly more) Blackwell informative policy than KG to persuade Receiver to take his preferred action. Equivalently, if experiments were costly to produce, as in Gentzkow and Kamenica (2014), then Sender would always need to consume less (resp. more) resources to persuade a wishful Receiver to take his preferred action than a Bayesian. The hypothesis of a binary state space facilitates the comparisons between the Bayesian-optimal and the wishful-optimal information policies as it ensures that the Bayesian-optimal and the wishful-optimal information policies are Blackwell comparable. Although the informativeness comparisons in corollary 5 do not necessarily extend when the state space contains more than two elements, Sender's welfare comparisons, in contrast, still hold under any arbitrary finite state space. We compare in figure 2.3 Sender's optimal information policies when Receiver is Bayesian and wishful, with the same payoff function as in figure 2.1. When the state space is finite, a policy $\tau \in \mathcal{T}(\mu_0)$ such that all elements in $\text{supp}(\tau)$ are affinely independent is (weakly) more Blackwell-informative than a policy $\tau' \in \mathcal{T}(\mu_0)$ if, and only if, and $\text{supp}(\tau') \subset \text{co}(\text{supp}(\tau))$ (see Lipnowski et al., 2020, Lemma 2). The support of the Bayesian-optimal policy τ^B (resp. wishful-optimal policy τ^W) is $\{\mu_-^B, \mu_{0,2}^B\}$ (resp. $\{\mu_-^W, \mu_{0,2}^W\}$). Hence, $\text{co}(\text{supp}(\tau^W)) = \{\mu \in \Delta(\Theta) : \exists t \in [0,1], \mu = t\mu_-^W + (1-t)\mu_{0,2}^W\}$. It is visible on figure 2.3 that $\{\mu_-^B, \mu_{0,2}^B\} \not\subset \text{co}(\text{supp}(\tau^W))$. Hence, τ^B and τ^W are not

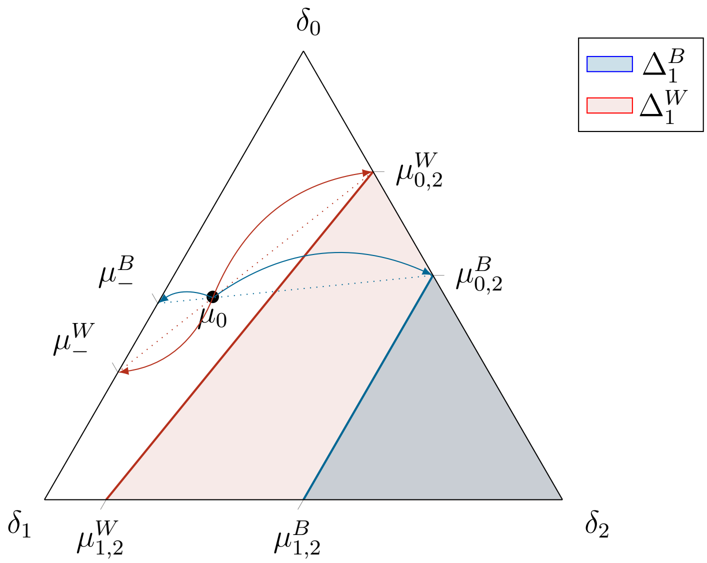


Figure 2.3: The Bayesian-optimal policy τ^B (in blue) vs. the wishful-optimal policy τ^W (in red) with respective supports $\{\mu_-^B, \mu_{0,2}^B\}$ and $\{\mu_-^W, \mu_{0,2}^W\}$.

Blackwell comparable. However, since Sender is interested in inducing action $a = 1$ and Receiver's favors that action, Sender's expected payoff is higher for any prior when Receiver is wishful.

2.5 Applications

In this section, we expose in three applications that corollary 5 might have important economic consequences.

2.5.1 Information provision and preventive health care

A public health agency (Sender) informs an individual (Receiver) about the prevalence of a certain disease. Receiver forms beliefs about the infection risk, which can be either high or low: $0 < \underline{\theta} < \overline{\theta} < 1$. The probability of contracting that illness also depends on whether the individual adopts a preventive treatment or not, where $a = 1$ designates adoption. Investment in the treatment entails a cost $c > 0$ to Receiver.[17] Moreover, let us assume that the effectiveness of the treatment, i.e., the probability that the treatment works, is $\alpha \in [0, 1]$ so that the probability of falling ill, conditional on adoption, is $(1 - \alpha)\theta$. The payoff from staying healthy is normalized to 0 whereas the payoff from being infected equals $-\varsigma < 0$ where ς is the severity of the disease.

[17] One might interpret that cost to be the price of the treatment or the either material or psychological cost from undertaking medical procedures.

Receiver's payoff function is

$$u(a,\theta) = (1-a)(-\varsigma\theta) + a(-(1-\alpha)\theta\varsigma - c)$$

for any $(a,\theta) \in A \times \Theta$. We assume that $\varsigma\alpha\underline{\theta} < c < \varsigma\alpha\overline{\theta}$ so Receiver faces a trade-off: he would prefer not to invest if he was sure the probability of infection was low and, conversely, would prefer to invest in the treatment if he was sure the risk of infection is high. Also remark that Receiver always expects to experience a negative payoff, as $u(a,\theta) < 0$ for any $(a,\theta) \in A \times \Theta$.

The public health agency wants to maximize the probability of individuals adopting the preventive treatment.[18] The agency informs individuals about the prevalence of the disease by designing and committing to a Bayes-plausible information policy τ. A Bayesian Receiver would be indifferent between adopting or not the treatment at belief

$$\mu^B = \frac{c - \alpha\underline{\theta}\varsigma}{\alpha(\overline{\theta} - \underline{\theta})\varsigma}.$$

In contrast, by proposition 4 and corollary 4, the equilibrium beliefs and behavior of a wishful Receiver are given by

$$\eta(\mu) = \begin{cases} \dfrac{\mu}{\mu + (1-\mu)\exp(\rho\varsigma(\overline{\theta} - \underline{\theta}))} & \text{if} \quad \mu < \mu^W \\[2ex] \dfrac{\mu\exp(-\rho(1-\alpha)\varsigma(\overline{\theta} - \underline{\theta}))}{\mu\exp(-\rho(1-\alpha)\varsigma(\overline{\theta} - \underline{\theta})) + (1-\mu)} & \text{if} \quad \mu \geq \mu^W \end{cases},$$

and

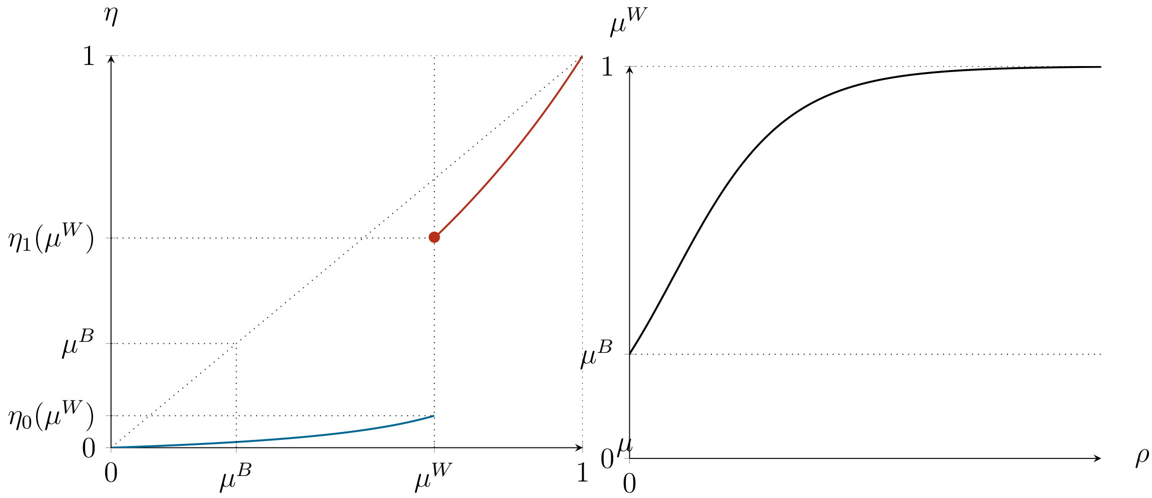
$$a(\eta(\mu)) = \mathbb{1}\left\{\mu \geq \mu^W\right\}$$

for any posterior belief $\mu \in [0,1]$, where

$$\mu^W = \frac{\exp(-\rho\underline{\theta}\varsigma) - \exp(\rho(-(1-\alpha)\underline{\theta}\varsigma - c))}{\exp(-\rho\varsigma\underline{\theta}) - \exp(\rho(-(1-\alpha)\underline{\theta}\varsigma - c)) + \exp(\rho(-(1-\alpha)\overline{\theta}\varsigma - c)) - \exp(-\rho\overline{\theta}\varsigma)}.$$

We illustrate the belief distortion of Receiver in figure 2.4a. Receiver is always overoptimistic about his probability of staying healthy, as $\eta(\mu) \leq \mu$ for any $\mu \in [0,1]$. Remark that non-adoption is associated with the highest possible payoff $-\varsigma\underline{\theta}$ as well as the highest payoff variability $\varsigma(\overline{\theta} - \underline{\theta})$. Accordingly, by lemma 3, Receiver always

[18] Maximizing the probability of adoption is a sensible objective since most infections cause negative externalities due to their transmission through social interactions. Therefore, a benevolent planner who wants to reduce the likelihood of transmission of an infection would do well to maximize the rate of adoption of the preventive treatment (for example, maximize condom distribution to control AIDS transmission, maximize injection of vaccines to control viral infections, or maximize mask use to control the spread of airborne diseases).



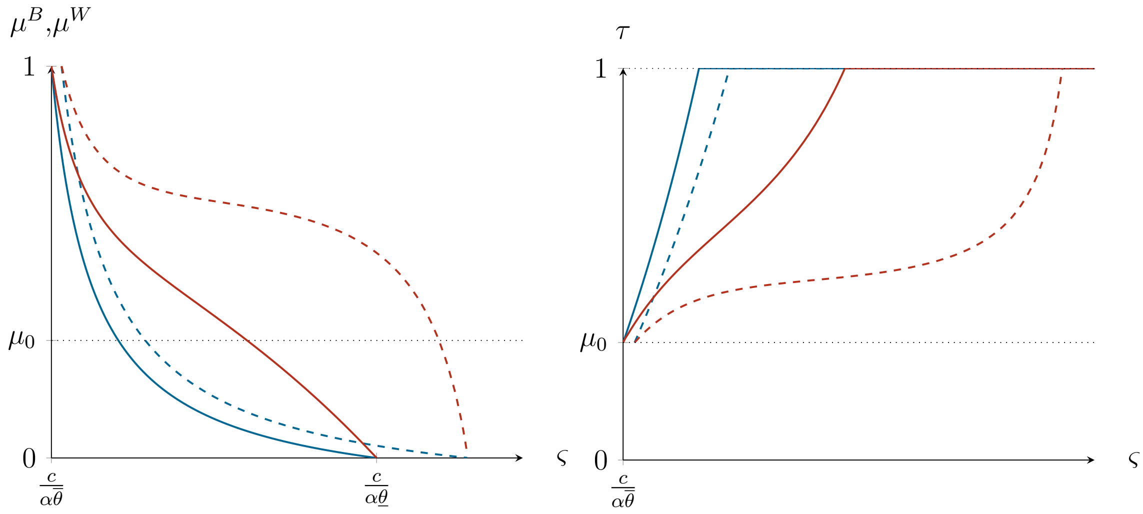
(a) equilibrium belief $\eta(\mu)$ as a function of μ. (b) Behavioral threshold μ^W as a function of ρ.

Figure 2.4: The belief correspondence for $\varsigma = 2$, $c = 0.5$, $\alpha = 0.8$, $\underline{\theta} = 0.1$, $\overline{\theta} = 0.9$ and $\rho = 2$. Receiver is always overoptimistic concerning his health risk for any induced posterior, except at $\mu = 0$ or $\mu = 1$. Moreover, the belief threshold μ^W as a function of ρ is strictly increasing and admits μ^B as a lower bound.

favors non adoption as illustrates figure 2.4b. As a result of corollary 5, Sender always needs to induce higher beliefs for Receiver to adopt the treatment than she would need if she faced a Bayesian agent, all the more so when Receiver's wishfulness ρ becomes larger. Therefore in this example, overoptimism of Receiver always goes against Sender's interest.

It is interesting to see how Sender's probability of inducing the adoption of the treatment evolves with respect to the severity of the disease ς, as well as the effectiveness of the treatment α.[19] We represent on figure 2.5b the probability that Sender induces adoption of the treatment under the optimal information policy as a function of ς. Notice that the probability of inducing adoption is less sensitive to the severity of the disease, i.e., becomes "flatter," when facing a wishful Receiver compared to the Bayesian when the treatment becomes less effective. The intuition is the following: when the treatment is fully effective, i.e., $\alpha = 1$, Receiver's payoff in case he invests in the treatment becomes state independent. Therefore, he does not have any incentive to distort beliefs when taking action $a = 1$. As a result, μ^W decreases and Receiver holds perfectly Bayesian beliefs when $\mu \geq \mu^W$. However, whenever there is uncertainty about the treatment efficacy, i.e., $\alpha < 1$, uncertainty about infection risk matters and gives room to belief distortion even when taking

[19]This probability is pinned down by the Bayes-plausibility constraint and equal to $\tau^{KG} = \mu_0/\mu^B$ in the Bayesian case and $\tau = \mu_0/\mu^W$ in the wishful case.



(a) Behavioral thresholds μ^B (in blue) and μ^W (in red) as functions of severity ς.

(b) Probability τ of inducing treatment adoption as a function of severity ς.

Figure 2.5: Red (resp. blue) curves correspond to wishful (resp. Bayesian) Receiver. We set parameters to $c = 0.5$, $\alpha = 0.8$, $\underline{\theta} = 0.1$, $\overline{\theta} = 0.9$ and $\rho = 2$. Full lines correspond to the case where $\alpha = 1$ whereas dashed curves correspond to $\alpha = 0.8$.

the treatment. Decreasing α increases the anticipated anxiety of Receiver leading to more optimistically biased beliefs, a higher μ^W and, in turn, complicates persuasion for Sender for any severity s. Remark on figure 2.5b that τ decreases sharply with α for a fixed s. In fact, one could show that as α decreases, τ becomes closer and closer to μ_0 for any ς, meaning that the agency cannot achieve a substantially higher payoff than under full disclosure.[20]

In the next subsection we extend out framework to the case of a continuous state space and linear preferences. We show that results in the finite state space case extend to this setting. We also highlight why we might expect persuasion to be more effective in the context of risky investment decisions.

2.5.2 Persuading a wishful investor

A financial broker (Sender) designs reports about the return of some risky financial product to inform a potential client (Receiver). The return of the product is $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$, where $\underline{\theta} < 0 < \overline{\theta}$. Returns are distributed according to the prior

[20] One additional implication of this result is the following. Assume the true treatment efficacy is α but Receiver perceives the efficacy to be $\hat{\alpha} < \alpha$ (e.g. because Receiver adheres to anti-vaccines movements or generally mistrusts the pharmaceutical industry). In that case, the doubts expressed by Receiver about the treatment efficacy makes him even more anxious which, in turn, makes belief distortion stronger and, thus, downplays the effectiveness of the agency's information policy whatever is the severity of the disease.

distribution μ_0. Let F be the cumulative distribution function associated with μ_0 and let us assume that μ_0 admits a continuous and strictly positive density function f over $[\underline{\theta}, \overline{\theta}]$. Receiver has some saved up money he is willing to invest and chooses action $a \in A = \{0, 1\}$, where $a = 0$ represents the choice of non-investing in which case Receiver's payoff is 0 and $a = 1$ represents investing, in which case Receiver's payoff is the realized return θ. The broker is remunerated on the basis of a flat fee $v > 0$ that is independent of the true product's profitability. Hence, Receiver's payoff is $u(a, \theta) = a\theta$ while Sender's payoff is $v(a, \theta) = va$ for any $(a, \theta) \in A \times \Theta$.

Receiver forms motivated beliefs about the return of the financial product. By proposition 4 his equilibrium beliefs are given by

$$\eta(\mu)(\tilde{\Theta}) = \begin{cases} \mu(\tilde{\Theta}) & \text{if } \displaystyle\int_\Theta \exp(\rho\theta)\, \mu(\mathrm{d}\theta) < 1 \\ \dfrac{\displaystyle\int_{\tilde{\Theta}} \exp(\rho\theta)\, \mu(\mathrm{d}\theta)}{\displaystyle\int_\Theta \exp(\rho\theta)\, \mu(\mathrm{d}\theta)} & \text{if } \displaystyle\int_\Theta \exp(\rho\theta)\, \mu(\mathrm{d}\theta) \geq 1 \end{cases},$$

for any $\mu \in \Delta(\Theta)$ and any Borel set $\tilde{\Theta} \subseteq \Theta$, and, by corollary 4, his equilibrium behavior is given by

$$a(\eta(\mu)) = \mathbb{1}\left\{\int_\Theta \exp(\rho\theta)\, \mu(\mathrm{d}\theta) \geq 1\right\}.$$

Therefore, Sender's indirect utility is equal to

$$v(\mu) = v\mathbb{1}\left\{\int_\Theta \exp(\rho\theta)\, \mu(\mathrm{d}\theta) \geq 1\right\}.$$

for any $\mu \in \Delta(\Theta)$. To make the problem interesting, we assume that neither a Bayesian nor a wishful Receiver would take action $a = 0$ under the prior. That is, $\hat{m} = \int_{\underline{\theta}}^{\overline{\theta}} \theta\mu_0(\mathrm{d}\theta) < 0$ and $\hat{x} = \int_{\underline{\theta}}^{\overline{\theta}} \exp(\rho\theta)\mu_0(\mathrm{d}\theta) < 1$.[21]

Under these assumptions, remark that a signal structure σ that induces a distribution τ over posterior beliefs μ matters for Receiver and Sender only through the *distribution of exponential moments* $x = \int_\Theta \exp(\rho\theta)\, \mu(\mathrm{d}\theta)$ *it induces.* Let X be the space of such moments, that is, $X = \mathrm{co}(\exp(\rho\Theta))$, where $\exp(\rho\Theta)$ is the graph of the function $\theta \mapsto \exp(\rho\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. That is, $X = [\underline{x}, \overline{x}]$ where $\underline{x} = \exp(\rho\underline{\theta})$ and $\overline{x} = \exp(\rho\overline{\theta})$. Let G be the prior cumulative distribution function over the random

[21]It is in fact always true that $\hat{m} < 0$ when $\hat{x} < 1$. Hence, assuming $\hat{m} < 0$ additionally to $\hat{x} < 1$ is without loss.

variable $\exp(\rho\theta)$ induced by F, that is

$$G(x) = F\left(\frac{\ln(x)}{\rho}\right),$$

for any $x \in [\underline{x}, \overline{x}]$. By standard arguments (Gentzkow and Kamenica, 2016), the problem of finding an optimal signal structure σ reduces to finding a cumulative distribution function H that maximizes

$$\int_{\underline{x}}^{\overline{x}} v(x)\,\mathrm{d}H(x)$$

subject to

$$\int_{\underline{x}}^{z} H(x)\,\mathrm{d}x \leq \int_{\underline{x}}^{z} G(x)\,\mathrm{d}x$$

for every $z \in [\underline{x}, \overline{x}]$. The solution to such a problem is well-known and can be found either using techniques from optimization under stochastic dominance constraints (Gentzkow and Kamenica, 2016; Ivanov, 2020; Kleiner et al., 2021) or linear programming (Kolotilin, 2018; Dworczak and Martini, 2019; Dizdar and Kováč, 2020). In our context, the optimal signal is a binary partition of the state space. That is, the broker reveals whether the return is above or below some threshold state.

Proposition 6. *There exists a unique $\theta^W \in [\underline{\theta}, \overline{\theta}]$ verifying*

$$\frac{1}{1 - F(\theta^W)} \int_{\theta^W}^{\overline{\theta}} \exp(\rho\theta) f(\theta)\,\mathrm{d}\theta = 1$$

and such that Sender pools all states $\theta \in [\theta^W, \overline{\theta}]$ under the same signal $s = 1$, i.e., $\sigma(1 \,|\, \theta) = 1$ for all $\theta \in [\theta^W, \overline{\theta}]$, and similarly pools all states $\theta \in [\underline{\theta}, \theta^W]$ under the same signal $s = 0$. Hence, the probability of inducing action $a = 1$ for Sender is equal to

$$\int_{\theta^W}^{\overline{\theta}} \sigma(1 \,|\, \theta) f(\theta)\,\mathrm{d}\theta = 1 - F(\theta^W).$$

Proof. See Ivanov (2020), Section 3. □

It is optimal for Sender to partition the state space at the threshold state making Receiver indifferent between investing or not at the prior. Such an information policy can intuitively be seen as the investment recommendation rule which maximizes the probability that Receiver invests given the prior distribution of returns F.

Using the exact same arguments as above, one can deduce that the probability of inducing action $a = 1$ when Receiver is Bayesian is given by $1 - F(\theta^B)$ where θ^B is

the unique threshold verifying the equation

$$\frac{1}{1-F(\theta^B)}\int_{\theta^B}^{\overline{\theta}} \theta f(\theta)\,\mathrm{d}\theta = 0.$$

Therefore, Sender is more effective at persuading a wishful Receiver if and only if $\theta^W < \theta^B$.

Proposition 7. *It is always true that* $\theta^W < \theta^B$. *Hence, Sender is always more effective at persuading a wishful rather than a Bayesian investor.*

Proof. See appendix B.6. □

The above result relates to proposition 5: buying the risky product is favored by the wishful investor since it is the action that yields both the highest possible payoff and the highest payoff variability. This example thus illustrates how the results in the finite state space case naturally extend to an infinite state space setting with linear preferences. It further helps explaining the pervasiveness of persuasion efforts in financial and betting markets, illustrating why some financial consulting firms seem to specialize in advice misconduct and cater to biased consumers.

2.5.3 Public persuasion and political polarization

A Sender (e.g., a politician, a lobbyist) persuades an odd-numbered finite group of voters $N = \{1, \ldots, n\}$ (e.g., a committee or parliamentary members) to adopt a proposal $x \in X = \{0, 1\}$, where $x = 0$ corresponds to the status-quo. The state space is binary, $\Theta = \{0, 1\}$, and the audience uses only the information disclosed by Sender to vote on the proposal. Let $a^i \in A = \{0, 1\}$ be the ballot cast by voter i, where $a^i = 0$ designates voting for the status-quo. The proposal is accepted if it is supported by a simple majority of voters. We assume Sender is only interested in the proposal being accepted, so her utility is $v(x) = x$. In contrast, any voter $i \in N$ has payoff function

$$u^i(x, \theta) = x\theta\beta^i + (1-x)(1-\theta)(1-\beta^i)$$

for any $(x, \theta) \in X \times \Theta$ where $\beta^i \in [0, 1]$ parametrizes the partisan preference of voter i. That is, all voters agree that the proposal should be implemented only when $\theta = 1$, but they vary in how much they value the implementation of the proposal. We assume β^i is symmetrically distributed around 1/2 in the population. Denote $\beta^m = 1/2$ the median voter's preference.

All voters form wishful beliefs and ρ is assumed homogeneous among the electorate. As a result, the direction as well as the magnitude of voters' belief distortion depends only on their partisan preferences β.[22] By proposition 4, voter i's belief under posterior $\mu \in [0,1]$ is given by

$$\eta(\mu,\beta^i) = \begin{cases} \dfrac{\mu}{\mu + (1-\mu)\exp(\rho(1-\beta^i))} & \text{if} \quad \mu < \mu^W(\beta^i) \\[2ex] \dfrac{\mu\exp(\rho\beta^i)}{\mu\exp(\rho\beta^i) + (1-\mu)} & \text{if} \quad \mu \geq \mu^W(\beta^i) \end{cases}.$$

where

$$\mu^W(\beta^i) = \frac{\exp(\rho(1-\beta^i)) - 1}{\exp(\rho(1-\beta^i)) + \exp(\rho\beta^i) - 2}.$$

Remark that, similarly as in Alonso and Câmara (2016), since the policy space is binary and voters do not hold private information there is no room for strategic voting in our model. Hence, citizen i's voting strategy under belief $\eta(\mu,\beta^i)$ is given by

$$a(\eta(\mu,\beta^i)) = \mathbb{1}\left\{\mu \geq \mu^W(\beta^i)\right\}.$$

Due to the heterogeneity in β, there is always some level of belief polarization among wishful voters for any $\mu \in]0,1[$. Let us measure such polarization by the sum of the absolute difference between each pair of beliefs in the audience

$$\pi(\mu) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} |\eta(\mu,\beta^i) - \eta(\mu,\beta^j)| \tag{2.6}$$

for any $\mu \in [0,1]$.

Proposition 8. *Under Sender's optimal information policy, the signal that leads to the implementation of the proposal also generates the maximum polarization among voters.*

Proof. See appendix B.5. □

To build an intuition of why this is the case, let's first note that, in our model, belief polarization and action polarization are closely related. Agents voting for the implementation of the proposal distort their beliefs upwards, whereas agents voting for the status quo distort their beliefs downwards. We can thus see that maximum belief polarization should be attained for some belief for which action polarization

[22]It has been shown in psychology (Babad et al., 1992; Babad, 1995, 1997) as well as in behavioral economics (Thaler, 2020) that voters political beliefs are often motivated by their partisan orientation.

is maximized, that is, for some belief at which $(n+1)/2$ agents are voting one way and the remaining $(n-1)/2$ are voting another way. This is the case for any $\mu \in [\mu^W(\beta^{m-1}), \mu^W(\beta^{m+1})[$.

Due to sincere voting, the result of the election always coincides with the vote of the median voter under posterior belief μ. Accordingly, Sender's indirect utility is

$$v(\mu) = \mathbb{1}\left\{\mu \geq \mu^W(\beta^m)\right\},$$

for any $\mu \in [0,1]$. The optimal information policy for Sender is thus supported on $\{0, \mu^W(\beta^m)\}$ whenever $\mu_0 \in]0, 1/2[$, and on $\{\mu_0\}$ whenever $\mu_0 \in]\mu^W(\beta^m), 1[$. The posterior $\mu^W(\beta^m)$, which leads to the implementation of the proposal, belongs to the interval $[\mu^W(\beta^{m-1}), \mu^W(\beta^{m+1})[$ and, as such, is in the neighbourhood of the belief that maximizes polarization for any distribution of preferences. When such distribution is symmetric around the median voter, polarization is maximized exactly at the middle point in that interval, which is $\mu^W(\beta^m)$.

We illustrate proposition 8 below in section 2.5.3 in a setup with 3 voters. Following corollary 4, wishful thinking induces voters to switch from disapproval to

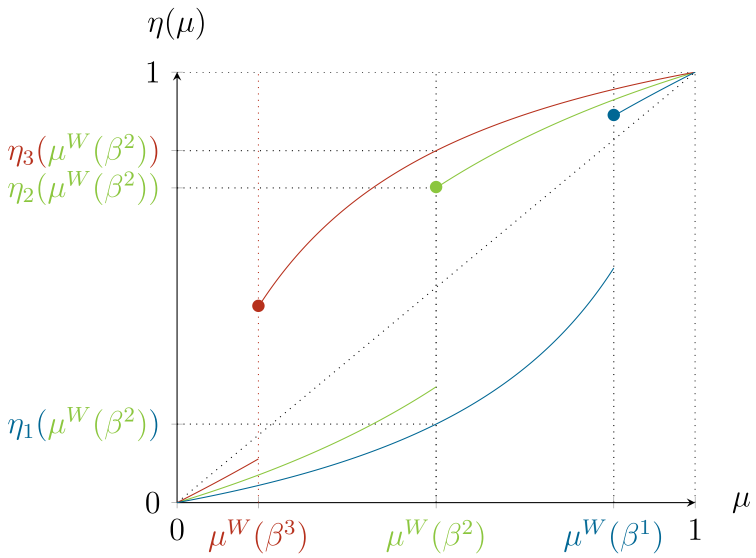


Figure 2.6: Beliefs distortions in the electorate for $\rho = 2$, $\beta_1 = 1/4$, $\beta_2 = 1/2$ and $\beta_3 = 3/4$. Polarization equals $\pi(\mu) = 2(\eta(\mu, \beta^1) - \eta(\mu, \beta^3))$ which is maximized at $\mu^W(\beta^2) = 1/2$.

approval at different Bayesian posteriors $\mu^W(\beta^i)$. The optimal information policy τ for Sender is the one that maximizes the probability of the median voter voting for the approval. That is, $\text{supp}(\tau) = \{0, \mu^W(\beta^m)\}$ and $\mu^W(\beta^m) = 1/2$ is induced with probability $\tau = \mu^W(\beta^m)/\mu_0$ whenever $\mu_0 \in]0, \mu^W(\beta^2)[$ and $\text{supp}(\tau) = \{\mu_0\}$ whenever $\mu_0 \in]\mu^W(\beta^2), 1[$.

Let us now turn to polarization. First, it is quite easy to see in section 2.5.3 that

$$\pi(\mu) = 2\left(\eta(\mu, \beta^1) - \eta(\mu, \beta^3)\right)$$

for any $\mu \in [0, 1]$, as the distances to the median belief add up to $\eta(\mu, \beta^1) - \eta(\mu, \beta^3)$. Thus, it suffices to check where $\eta(\mu, \beta^1) - \eta(\mu, \beta^3)$ is maximized. Quite naturally, polarization is maximized when the posterior belief induced by Sender is in between $\mu^W(\beta^3)$ and $\mu^W(\beta^1)$. In particular, it is exactly maximized at the posterior belief $\mu^W(\beta^2) = 1/2$ which is exactly the posterior belief Sender induces to obtain the approval of the proposal under her optimal policy.

proposition 8 establishes that the intuition developed in this example is generally valid when the partisan preferences of voters are symmetrically distributed around the median. In other words, attempts by a rational sender to maximize the probability of approval induces, as an externality, maximal belief polarization among wishful voters. This result differs from the literature studying the possible heterogeneity of beliefs due to deliberate attempts at persuasion which tends to focus on polarization arising from differential access to information.[23] Our model gives an alternative mechanism to the rise of polarization, based on motivated beliefs: a sender can induce polarization involuntarily when her message is subject to motivated interpretations, and such polarization might be especially large whenever sender's strategy involves targeting an agent with a median preference.

2.6 Conclusion

In this paper we study optimal persuasion in the presence of a wishful Receiver. By modeling wishful thinking as a process that optimally trades-off gains in anticipatory utility with the cost of distorting beliefs, we characterize the correspondence between wishful and Bayesian beliefs, highlighting the particularities that such belief formation process entails.

In particular, we show that wishful thinking impacts behavior, causing some actions to be favored in the sense that they are taken at a greater set of beliefs. This has important implications for the strategic design of information, as it adds some nuance on the way preferences and information determine behavior. Concretely, we show that, in the presence of wishful thinking, persuasion is more effective when it is aimed at inducing actions that are risky but can potentially yield a very large payoff and less effective when it is aimed at inducing more cautious actions. We

[23] See Arieli and Babichenko (2019) for general considerations on the private persuasion of multiple receivers and see Chan et al. (2019) for an application to voting.

use this model to illustrate why information disclosure seems less effective than expected at inducing preventive health behavior and more effective than expected at inducing dubious financial investments. Wishful thinking opens a channel for preferences to interfere in belief formation, raising the question of what kind of belief polarization could we observe in a population in which agents have access to the same information but vary in their preferences. We show in an application that an information designer interested in the approval of a proposal would, by optimally targeting the median voter in her choice of signal structure, induce, as an externality, maximum polarization among the electorate whenever the proposal is approved.

Some studies already investigate the effects of wishful thinking on the outcomes of strategic interactions (see, Yildiz, 2007; Banerjee et al., 2020; Heller and Winter, 2020). Further investigation on ways in which individual preferences might impact information processing and how these may impact social phenomena such as belief polarization in non-strategic and strategic settings seem to be promising paths for future research.

Bibliography

Abeler, J., Becker, A., and Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, 113:96–104.

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, 87(4):1115–1153.

Akerlof, G. A. and Dickens, W. T. (1982). The Economic Consequences of Cognitive Dissonance. *American Economic Review*, 72(3):307–319.

Alonso, R. and Câmara, O. (2016). Persuading Voters. *American Economic Review*, 106(11):3590–3605.

Arieli, I. and Babichenko, Y. (2019). Private Bayesian persuasion. *Journal of Economic Theory*, 182:185–217.

Babad, E. (1995). Can Accurate Knowledge Reduce Wishful Thinking in Voters' Predictions of Election Outcomes? *The Journal of Psychology*, 129(3):285–300.

Babad, E. (1997). Wishful thinking among voters: motivational and cognitive Iinfluences. *International Journal of Public Opinion Research*, 9(2):105–125.

Babad, E., Hills, M., and O'Driscoll, M. (1992). Factors Influencing Wishful Thinking and Predictions of Election Outcomes. *Basic and Applied Social Psychology*, 13(4):461–476.

Babad, E. and Katz, Y. (1991). Wishful Thinking—Against All Odds. *Journal of Applied Social Psychology*, 21(23):1921–1938.

Banerjee, S., Davis, J., and Gondhi, N. (2020). Motivated Beliefs in Coordination Games. *SSRN Electronic Journal*.

Bénabou, R. (2015). The Economics of Motivated Beliefs. *Revue d'économie politique*, 125(5):665–685.

Bénabou, R. and Tirole, J. (2002). Self-Confidence and Personal Motivation. *Quarterly Journal of Economics*, 117(3):871–915.

Bénabou, R. and Tirole, J. (2004). Willpower and Personal Rules. *Journal of Political Economy*, 112(4):848–886.

Bénabou, R. and Tirole, J. (2006). Belief in a Just World and Redistributive Politics. *Quarterly Journal of Economics*, 121(2):699–746.

Bénabou, R. and Tirole, J. (2011). Identity, Morals, and Taboos: Beliefs as Assets *. *Quarterly Journal of Economics*, 126(2):805–855.

Bénabou, R. and Tirole, J. (2016). Mindful Economics: The Production, Consumption, and Value of Beliefs. *Journal of Economic Perspectives*, 30(3):141–164.

Benjamin, D., Bodoh-Creed, A., and Rabin, M. (2019). Base-Rate Neglect: Foundations and Implications.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics: Applications and Foundations 2*, volume 2, chapter 2, pages 69–186. Elsevier B.V.

Bergemann, D. and Morris, S. (2016). Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium. *American Economic Review*, 106(5):586–591.

Bergemann, D. and Morris, S. (2019). Information Design: A Unified Perspective. *Journal of Economic Literature*, 57(1):44–95.

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2018). Behavioral Household Finance. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics: Applications and Foundations 1*, chapter 3, pages 177–276. Elsevier B.V.

Bracha, A. and Brown, D. J. (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, 75(1):67–80.

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal Expectations. *American Economic Review*, 95(4):1092–1118.

Caplin, A. and Leahy, J. (2001). Psychological Expected Utility Theory and Anticipatory Feelings. *Quarterly Journal of Economics*, 116(1):55–79.

Caplin, A. and Leahy, J. (2019). Wishful Thinking. *NBER Working Paper Series*.

Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E., and Crockett, M. J. (2020). Motivated misremembering of selfish decisions. *Nature Communications*, 11(1):2100.

Chan, J., Gupta, S., Li, F., and Wang, Y. (2019). Pivotal persuasion. *Journal of Economic Theory*, 180:178–202.

Chandra, A., Handel, B., and Schwartzstein, J. (2019). Behavioral economics and health-care markets. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics: Applications and Foundations 2*, chapter 6, pages 459–502. Elsevier B.V.

Chew, S. H., Huang, W., and Zhao, X. (2020). Motivated False Memory. *Journal of Political Economy*, 128(10):3913–3939.

Coutts, A. (2019). Testing models of belief bias: An experiment. *Games and Economic Behavior*, 113:549–565.

de Clippel, G. and Zhang, X. (2020). Non-Bayesian Persuasion. *Working Paper*.

Dizdar, D. and Kováč, E. (2020). A simple proof of strong duality in the linear persuasion problem. *Games and Economic Behavior*, 122:407–412.

Dupas, P. (2011). Health Behavior in Developing Countries. *Annual Review of Economics*, 3(1):425–449.

Dworczak, P. and Martini, G. (2019). The Simple Economics of Optimal Persuasion. *Journal of Political Economy*, 127(5):1993–2048.

Egan, M., Matvos, G., and Seru, A. (2019). The Market for Financial Adviser Misconduct. *Journal of Political Economy*, 127(1):233–295.

Eliaz, K., Spiegler, R., and Thysen, H. C. (2021a). Persuasion with endogenous misspecified beliefs. *European Economic Review*, 134:103712.

Eliaz, K., Spiegler, R., and Thysen, H. C. (2021b). Strategic interpretations. *Journal of Economic Theory*, 192:105192.

Engelmann, J., Lebreton, M., Schwardmann, P., van der Weele, J. J., and Chang, L.-A. (2019). Anticipatory Anxiety and Wishful Thinking. *SSRN Electronic Journal*.

Ettinger, D. and Jehiel, P. (2010). A Theory of Deception. *American Economic Journal: Microeconomics*, 2(1):1–20.

Exley, C. and Kessler, J. (2019). Motivated Errors. *NBER Working Paper Series.*

Eyster, E. (2019). Errors in strategic reasoning. In *Handbook of Behavioral Economics: Applications and Foundations 2*, volume 2, chapter 3, pages 187–259. Elsevier B.V.

Ganguly, A. and Tasoff, J. (2017). Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future. *Management Science*, 63(12):4037–4060.

Gentzkow, M. and Kamenica, E. (2014). Costly Persuasion. *American Economic Review: Papers & Proceedings*, 104(5):457–462.

Gentzkow, M. and Kamenica, E. (2016). A Rothschild-Stiglitz Approach to Bayesian Persuasion. *American Economic Review: Papers & Proceedings*, 106(5):597–601.

Golman, R., Hagmann, D., and Loewenstein, G. (2017). Information Avoidance. *Journal of Economic Literature*, 55(1):96–135.

Golman, R., Loewenstein, G., Moene, K. O., and Zarri, L. (2016). The Preference for Belief Consonance. *Journal of Economic Perspectives*, 30(3):165–188.

Hagenbach, J. and Koessler, F. (2020). Cheap talk with coarse understanding. *Games and Economic Behavior*, 124:105–121.

Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton University Press.

Heger, S. A. and Papageorge, N. W. (2018). We should totally open a restaurant: How optimism and overconfidence affect beliefs. *Journal of Economic Psychology*, 67(July):177–190.

Heller, Y. and Winter, E. (2020). Biased-Belief Equilibrium. *American Economic Journal: Microeconomics*, 12(2):1–40.

Ivanov, M. (2020). Optimal monotone signals in Bayesian persuasion mechanisms. *Economic Theory.*

Jiao, P. (2020). Payoff-Based Belief Distortion. *The Economic Journal*, 130(629):1416–1444.

Kamenica, E. (2019). Bayesian Persuasion and Information Design. *Annual Review of Economics*, 11:249–272.

Kamenica, E. and Gentzkow, M. (2011). Bayesian Persuasion. *American Economic Review*, 101(6):2590–2615.

Kleiner, A., Moldovanu, B., and Strack, P. (2021). Extreme Points and Majorization: Economic Applications. *Econometrica*, 89(4):1557–1593.

Kolotilin, A. (2018). Optimal information disclosure: A linear programming approach. *Theoretical Economics*, 13(2):607–635.

Kremer, M., Rao, G., and Schilbach, F. (2019). Behavioral development economics. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics: Applications and Foundations 2*, chapter 5, pages 345–458. Elsevier B.V.

Krizan, Z. and Windschitl, P. D. (2009). Wishful Thinking about the Future: Does Desire Impact Optimism? *Social and Personality Psychology Compass*, 3(3):227–243.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, 53(4):636–647.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3):480–498.

Le Yaouanq, Y. (2021). Motivated cognition in a model of voting. *Working Paper*.

Lerman, C., Hughes, C., Lemon, S. J., Main, D., Snyder, C., Durham, C., Narod, S., and Lynch, H. T. (1998). What you don't know can hurt you: adverse psychologic effects in members of BRCA1-linked and BRCA2-linked families who decline genetic testing. *Journal of Clinical Oncology*, 16(5):1650–1654.

Levy, G., Moreno de Barreda, I., and Razin, R. (2018). Persuasion with Correlation Neglect. *Working Paper*.

Lipnowski, E., Mathevet, L., and Wei, D. (2020). Attention Management. *American Economic Review: Insights*, 2(1):17–32.

Loewenstein, G. (1987). Anticipation and the Valuation of Delayed Consumption. *The Economic Journal*, 97(387):666–684.

Mayraz, G. (2011). Wishful Thinking. *SSRN Electronic Journal*.

Mijović-Prelec, D. and Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1538):227–240.

Mullainathan, S., Noeth, M., and Schoar, A. (2012). The Market for Financial Advice: An Audit Study. Technical report, National Bureau of Economic Research, Cambridge, MA.

Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse Thinking and Persuasion *. *Quarterly Journal of Economics*, 123(2):577–619.

Oster, E., Shoulson, I., and Dorsey, E. R. (2013). Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease. *American Economic Review*, 103(2):804–830.

Saucet, C. and Villeval, M. C. (2019). Motivated memory in dictator games. *Games and Economic Behavior*, 117:250–275.

Schwardmann, P. (2019). Motivated health risk denial and preventative health care investments. *Journal of Health Economics*, 65:78–92.

Strzalecki, T. (2011). Axiomatic Foundations of Multiplier Preferences. *Econometrica*, 79(1):47–73.

Thaler, M. (2020). The 'Fake News' Effect: Experimentally Identifying Motivated Reasoning Using Trust in News. *SSRN Electronic Journal*.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5):806–820.

Yildiz, M. (2007). Wishful Thinking in Strategic Environments. *Review of Economic Studies*, 74(1):319–344.

Chapter 3

Text and Subtext

Abstract

We study a persuasion problem in which a sender faces an audience that is heterogeneous both in their preferences and in the extent to which they understand messages. The sender is able to exploit such heterogeneity to convey some information privately to some receivers – the subtext –, but is constrained by the publicly understood aspects of its own communication strategy – the text –. We characterize the set of joint distributions of posteriors that the sender can feasibly induce and show that the sender's value from the problem can be retrieved through a recursive concavification procedure.

JEL classification codes: D82, D83, D90.
Keywords: Information design; persuasion; language; bounded rationality.

[0]We thank Victor Augias, Jeanne Hagenbach and Eduardo Perez for the helpful discussions, as well as seminar audiences in Sciences Po.

3.1 Introduction

> *Metaphysics should be written with accurate definitions and demonstrations, but nothing should be demonstrated in it that conflicts too much with received opinions. For thus this metaphysics will be able to be received. If it is once approved, then afterwards, if any examine it more profoundly, they will draw the necessary consequences themselves.*
>
> —Gottfried Wilhelm Leibniz[1]

In many instances of economic and political life, communication with a plurality of receivers is neither purely public nor purely private: some aspects of the information transmitted might be commonly understood – what we refer to as the *text* –, whereas finer aspects – the *subtext* – might only be observed by a subset of the audience. These settings allow for a mixed mode of communication, one that is more permissive than public communication, as it allows for some information to be transmitted privately through the subtext, but more restrictive than private communication, since one cannot target privately the receivers who only have access to the text.

Think for instance of a hierarchical organization, where messages sent to lower ranks of the organization might also be observed by the upper ranks, whereas messages sent to upper ranks are not observed by those in lower echelons. In this case communication with the members of the organization exhibits the feature of varying degrees of refinement along the organization's ranks: whereas members of the lowest echelon only have access to information that is public within the organization – the text –, members of higher ranks have varying degrees of additional information – the subtext –.

Another example of such mixed mode of communication is what is termed in politics as *dog-whistling*: the usage of coded language designed to signal something to some groups (those who recognise the term) without antagonizing others (those who don't). An example, taken from Albertson (2015), illustrates how such communication strategy might be used: In his 2003 State of the Union Address, George W. Bush declared that "there is power, wonder-working power, in the goodness and idealism and faith of the American people". While most of the listeners would not infer any particular meaning from such phrase, evangelical listeners could recognize the term "wonder-working power" from a popular hymn, and thus perceive in this term a signal for them. While appealing explicitly to evangelicals could alienate part of the

[1] Leibniz continues: "In this metaphysics, it will be useful for there to be added here and there the authoritative utterances of great men, who have reasoned in a similar way; especially when these utterances contain something that seems to have some possible relevance to the illustration of a view".

audience, doing so in a coded manner enabled Bush to convey some information privately to some members of the audience.

The aim of this paper is to study communication in multi-receiver settings where the audience – either due to the organizational structure within which they are embedded or due to heterogeneity in receiver's ability to decode messages – exhibits varying degrees of refinement with respect to the information they might have access to. We study a model in which a sender designs an information structure to persuade an audience to act in a certain way. Members of the audience vary in their preferences, but also in how finely they are able to extract information from the realized message. As in Blume and Board (2013), differences in refinement are modeled as differences in receivers ability to distinguish between different messages.

Such heterogeneity in refinement gives the sender leeway to convey some information privately through the subtext. What the sender can convey through the subtext, however, is constrained by what is conveyed through the text. In section 3.3.1 we characterize the joint distributions of posterior beliefs that can be induced by the sender: these are any joint distribution of posteriors such that i) the expected posterior of the coarsest receiver is equal to the prior and ii) conditional on the realization of a given posterior (call it μ) for some receiver, the expected posterior of any more refined receiver is equal to μ. We then show that the maximum payoff that the sender can achieve in the persuasion problem can be retrieved by a process of "recursive concavification", which is formally defined in section 3.3.2.

3.1.1 Related Literature

This paper relates to the literature exploring the role of limitations to communication in information transmission, and in particular on information design (see Bergemann and Morris (2019)).

Our setting is close to the one studied in Kamenica and Gentzkow (2011), who characterize optimal experiments under public communication. We expand their characterization of feasible distributions of posteriors and their concavification method to settings where communication is neither purely public nor purely private. Aybas and Turkel (2022) study a persuasion problem where the number of available messages is smaller than the number of states of the world or actions of the receiver, such that communication is inherently coarse. As in the present paper, they show that the value of persuasion is given by a modified concave envelope of the sender's indirect utility. This limitation in the set of available messages is also present in Le Treust and Tomala (2019).

Another related problem can be found in Bloedel and Segal (2018), who study

a persuasion problem with a rationally inattentive receiver. Like in our paper, coarseness in the receiver's understanding is central to their analysis, but their focus is on endogenizing such coarseness through attention whereas our focus is on the role of heterogeneity in such coarseness across different receivers. Other limitations on receiver's interpretation have also been explored by the literature (Eliaz et al., 2021; Levy et al., 2022; Schwartzstein and Sunderam, 2021).

The idea of limited language has also been used to study communication in several different contexts. Blume and Board (2013) study the role of limitations to language in the context of coordination games. Like the present paper, they model language competence as partitions of the set of available messages, although for most of their results they focus on a class of partitions that is more restricted than the one considered in the present paper. Hagenbach and Koessler (2020) study limited language competence on the part of the sender in cheap talk games.

Our paper also has a close dialogue with the literature on information design in networks (Galperti and Perego, 2020; Egorov and Sonin, 2020; Corrao, 2021; Liporace, 2021), in which receivers are embedded in a network describing how information "leaks" between one receiver and another. This raises questions regarding optimal seeding, privacy, and so on. The model present in this paper can be seen as one that analyses a particular network structure, in which information flows in a particular direction, resulting in receivers that can be ordered in terms of their information.

3.2 Model

3.2.1 Setup

A sender designs an information structure to persuade the members of an audience to act in a certain way. All relevant uncertainties are summarized by the state of the world ω belonging to a finite set Ω and all players have a common prior $\mu_0 \in \Delta(\Omega)$ with full support.

The audience is composed of n receivers $i \in \{1, \cdots, n\}$, with different receivers potentially differing in their preferences and their partial understanding of the messages. Receiver i has preferences $u_i : A_i \times \Omega \to \mathbb{R}$, where A_i is the set of actions from which the receiver chooses.

Sender's preferences are given by $v : A \to \mathbb{R}$, where $A = A_1 \times \cdots \times A_n$. We assume that v is additively separable in receiver's actions such that we can write $v(a_1, \cdots, a_n) = \sum_{i=1}^{n} v_i(a_i)$. Sender can design an information structure $\sigma : \Omega \to \Delta(M)$ to inform the audience, where M is a fixed set of messages.

3.2.2 Partial Understandings

Each receiver is endowed with an understanding, which defines whether such receiver is able to differentiate between any two messages $m, m' \in M$. Formally, a receiver i's understanding is a partition $P_i = \{p_i^1, \cdots, p_i^{k(i)}\}$, where P_i is a collection of nonempty disjoint subsets of M that completely cover M.

Receiver i is unable to distinguish between messages that belong to the same partition element p_i, and as such must update his beliefs in the same way following the realization of any such messages. Denote $p_i(m)$ the element of P_i that includes a message m. Following the realization of m, receiver i forms posterior belief:

$$\mu_i(\omega|p_i(m)) = \frac{\sigma(p_i(m)|\omega)\mu_0(\omega)}{\sigma(p_i(m))} = \frac{[\sum_{m' \in p_i(m)} \sigma(m'|\omega)]\mu_0(\omega)}{\sum_{m' \in p_i(m)} \sigma(m')}$$

As such, an audience's understanding is characterized by a collection of partitions $\{P_i\}_{i \in \{1, \cdots, n\}}$ as well as a collection of preferences $\{u_i\}_{i \in \{1, \cdots, n\}}$. In order to delimit the types of understandings we consider, we introduce two definitions:

Definition 5 (Partition refinement)**.** *A partition P' is a refinement of partition P if every element of P' is a subset of some element of P.*

Definition 6 (Refinement order)**.** *A collection of partitions $\{P_i\}_{i \in \{1, \cdots, n\}}$ is said to allow for a refinement order if P_j is a refinement of P_i whenever $i < j$.*

In this paper we consider collections of partitions satisfying a refinement order, such that we can label the different members of the audience according to how finely they are able to understand the informational content of messages. Whenever a message $m \in M$ is realized, $p_1(m)$ can be seen as the *text* - the aspect of the message that is commonly understood by all members of the audience -, whereas $p_i(m)$ for $i \geq 2$ represent the different depths of *subtext* present in the message.

3.2.3 Two Interpretations of the Model

There are two ways one might interpret the model. One is the literal interpretation that each message realization $m \in M$ is public, but agents vary in how finely they might understand its informational content. This interpretation relates to the notion of language competence developed in Blume and Board (2013). Under this interpretation, one could think of the set M as a set of sentences in English, for instance. A receiver who does not speak English at all won't be able to differentiate between any of the sentences and thus won't extract any information from a given

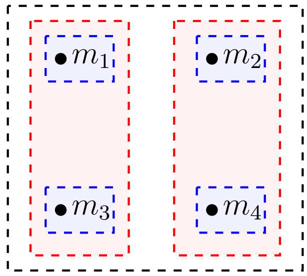


Figure 3.1: Three partitions satisfying a refinement order: In black P_1, in red P_2 and in blue P_3.

message, whereas a receiver with some knowledge of such language will be able to distinguish between more sentences and thus be able to capture finer meaning from it. Such differences are present even among native speakers: whereas some people might not distinguish between two words with the same denotative meaning (text), others might be aware of differences in their connotative meaning (subtext), and thus be responsive to the usage of one word rather than the other.

Importantly, such heterogeneity in understanding is often reflective of heterogeneity in group identity: people's ability at identifying something as particularly meaningful depends on their past experiences, education or interests, all of which tend to be correlated with their preferences.

A second interpretation, closer to the idea of information systems present in Galperti and Perego (2020), is that the sender is constrained in its ability to target different groups. Imagine for instance that the audience is composed of two receivers, where the sender is contrained at only communicating publicly with receiver 1 whereas receiver 2 can be targeted privately. This setting could be represented by a message space M containing different messages (m_1, m_2), where m_1 is the realization of the public message and m_2 the realization of the private message and where P_2 is a fully refined partition of M whereas P_1 is composed of several partition elements pooling together, for a given realisation of m_1, all the different possible realizations of m_2.

3.3 Results

3.3.1 Feasibility

In this section we characterize the feasible distributions of posteriors that the sender can induce through its strategy σ, as well as the value that she can achieve through persuasion.

Lemma 4. *Consider two receivers i and j such that P_j is a refinement of P_i. Then,*

$$\mu_i(\omega|p_i) = \sum_{p_j:p_j \subseteq p_i} \sigma(p_j|p_i)\mu_j(\omega|p_j)$$

Lemma 4 ties together the posterior beliefs of two agents j and i: it establishes that, following the realization of some message belonging to a partition element p_i of the least refined receiver, the expected posterior belief of receiver j must be i's realized posterior $\mu_i(\omega|p_i)$.

Sender's communication strategy σ induces a joint distribution of beliefs in the audience, which we denote by τ. The following proposition characterizes the joint distributions of posterior beliefs that the sender can feasibly induce given some strategy σ.

Proposition 9 (Feasible distributions of posteriors)**.** *Let the audience's understandings $\{P_i\}_{i\in\{1,\cdots,n\}}$ satisfy a refinement order. Sender can induce any joint distribution of posteriors $\tau(\mu_1, \cdots, \mu_n)$ such that:*

$$\sum_{Supp(\tau_1)} \mu_1\tau_1(\mu_1) = \mu_0$$

and

$$\sum_{Supp(\tau_{j|i})} \mu_j\tau_{j|i}(\mu_j|\mu_i) = \mu_i, \forall i < j.$$

Proposition 9 establishes that if the audience's understandings satisfy a refinement order, Sender's strategy σ can induce any joint distribution of posteriors such that i) the beliefs of the coarsest agent satisfy the standard Bayes Plausibility condition (Kamenica and Gentzkow, 2011) and ii) conditional on the realization of some posterior μ of some agent, the expected posterior of any more refined agent must be μ. Note that these conditions imply that the beliefs of any agent $i \in \{1, \cdots, n\}$ also satisfy the Bayes Pausibility condition.

3.3.2 Optimality

Proposition 9 establishes that the problem of the Sender can be viewed as a sequential information design problem: one of designing the distribution of beliefs of the coarsest agent and then, conditional on that, designing the distribution of beliefs of the second coarsest agent, and so forth.

Since v is additively separable on the actions taken by the audience, we can denote Sender's indirect utility as $\hat{v}(\mu_1, \cdots \mu_n) = \sum_{i=1}^{n} \hat{v}_i(\mu_i)$, where $\hat{v}_i(\mu_i) = v_i(\hat{a}_i(\mu_i))$.

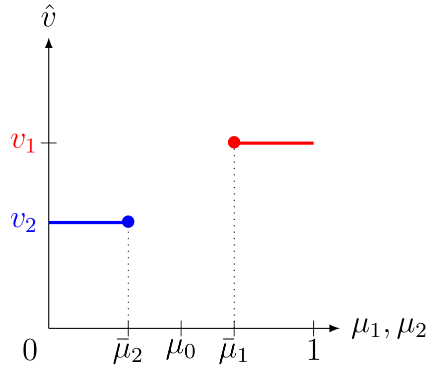


Figure 3.2: Sender's indirect utility

Define:

$$V_i(\mu_{i-1}) \equiv \begin{cases} \sup\{z|(\mu_{i-1}, z) \in \text{co}(\hat{v}_i + V_{i+1})\} & \text{for } i \in \{1, \cdots, n-1\} \\ \sup\{z|(\mu_{i-1}, z) \in \text{co}(\hat{v}_i)\} & \text{for } i = n \end{cases}$$

Proposition 10. *(Recursive Concavification) The value of an optimal signal for the sender is given by $V_1(\mu_0)$.*

The standard approach for finding the sender-optimal information structure in persuasion settings involves computing the concave envelope of sender's indirect utility function and evaluating it at the prior belief. This approach is suitable in the multi-receiver case when communication is entirely public (i.e. when there's only a text), but doesn't apply directly in our setting since here receivers of different groups will form a different posterior after observing the same message. Instead, in our case the sender-optimal information structure can be found recursively, by identifying what would be the optimal distribution of μ_n conditional on some realization of μ_{n-1} and then moving backwards and incorporating the value of such optimal distribution into the identification of the optimal distribution of μ_{n-1}, and so on.

3.4 Example

Consider a simple setting where the audience is composed of two agents $i \in \{1, 2\}$, each of whom chooses an action $a_i \in \{l, r\}$. Receiver's payoffs depend on their actions and on a binary state of the world $\omega \in \{L, R\}$, distributed according to a common prior $\mu_0 = Pr(\omega = R)$.

Sender wants to induce receivers to take $a = r$ and has utility $v(a_1, a_2) = v_1 \mathbb{1}\{a_1 = r\} + v_2 \mathbb{1}\{a_2 = r\}$. For that purpose she chooses among a family of distributions $\{\sigma(\cdot|\omega)\}_{\omega \in \{L,R\}}$ over a message space $M = \{m_1, m_2, m_3, m_4\}$.

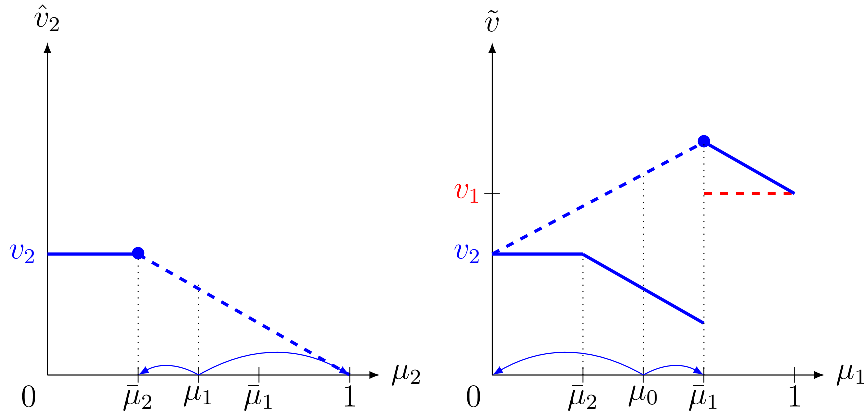


Figure 3.3: Recursive concavification

Each receiver has a distinct preference. Define $\Delta u_i^R = u_i(r, R) - u_i(l, R)$ and $\Delta u_i^L = u_i(r, L) - u_i(l, L)$, such that a each of the receivers has a distinct threshold belief $\bar{\mu}_i = -\frac{\Delta u_i^L}{\Delta u_i^R - \Delta u_i^L}$ under which they are indifferent between each of the actions. Consider $\Delta u_1^R, \Delta u_2^L > 0$ and $\Delta u_1^L, \Delta u_2^R < 0$, such that receiver 1 wants to match the state whereas receiver 2 does not, and assume that $\bar{\mu}_2 < \bar{\mu}_1$, such that at no belief both receivers would be willing to take $a = r$. Figure 3.2 illustrates sender's indirect utility $\hat{v}$ in this case.

Consider first the case where both receivers hold the same partition P, with $|P| \geq 2$. In that case there is no heterogeneity in receivers understandings, meaning that they will form the same posterior beliefs after the realization of any message realization. In this case the sender is never able to induce both agents to simultaneously take her preferred action, and as such designs σ so as to target one of the receivers optimally. Sender's value in this case is given by the concave envelope of $\hat{v}$ evaluated at the prior belief μ_0.

Now imagine that receivers differ in their understanding of the message, and instead hold the following partitions:

$$P_1 = \{\{m_1, m_2\}, \{m_3, m_4\}\}$$
$$P_2 = \{\{m_1\}, \{m_2\}, \{m_3\}, \{m_4\}\}$$

The sender can now exploit the audience's heterogeneous understandings in order to convey some information privately to receiver 2 by choosing different conditional distributions to messages that belong to the same partition element of receiver 1. This amounts to designing a subtext (for instance the informational content of m_1 and m_2) *conditional* on the text (the realization of $\{m_1, m_2\}$).

To understand what the sender can achieve through the subtext, consider the left panel of figure 3.3. From proposition 9 we know that, given any posterior realization

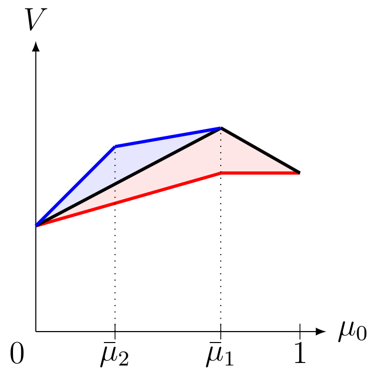


Figure 3.4: The value of persuasion under different modes of communication: text and subtext in Black, public in red and private in blue.

μ_1 of the coarsest receiver, the more refined receiver can hold any distribution of posteriors that average to μ_1. As such, for a given μ_1, sender can achieve through the subtext any value z such that $(\mu_1, z) \in \text{co}(\hat{\text{v}}_2)$. An optimal subtext is then given by the distribution of posteriors that achieve $\sup\{z|(\mu_1, z) \in \text{co}(\hat{v}_2)\}$, as illustrated in figure 3.3.

Knowing that the value of the subtext is given by the concave envelope of $\hat{v}_2$ allows us to know precisely the payoff that can be achieved through the text: for any belief μ_1 that sender generates, its value is given by the payoff it achieves from the coarsest agent *plus* the value of the subtext that is achievable under such μ_1. As such, by summing $\hat{v}_1$ with the concave envelope of $\hat{v}_2$ we obtain a function denoting the maximum payoff that the sender can achieve for any belief μ_1 that it induces. Taking the concave envelope of this function and evaluating it at μ_0 tells us the value that the sender can achieve given the prior.

Figure 3.4 shows a comparison of the value of persuasion under different modes of communication. It depicts in red the value under public communication, in blue the value under private communication and in black the value with text and subtext. A few things are worth noting: first, the value of both private communication and communication with text and subtext is everywhere above the value of public communication. This is because non-public modes of communication always have some probability of inducing both agents to simultaneously take the sender's preferred action. Second, the value of private communication is greater than the value of text and subtext for $\mu_0 < \bar{\mu}_1$, but both values coincide for $\mu_0 \geq \bar{\mu}_1$. This is so because when $\mu_0 \geq \bar{\mu}_1$ receiver 1 is already taking sender's preferred action by default, such that she can be quiet with the text and simply target receiver 2 optimally with the subtext. Whenever $\mu_0 < \bar{\mu}_1$, however, sender needs to convey some information to receiver 1 in order to persuade him to choose $a_1 = r$. In doing so she makes the

task of persuading receiver 2 harder, as both receivers have opposite preferences and require different information to be convinced.

Bibliography

Albertson, B. (2015). Dog-whistle politics: Multivocal communication and religious appeals. *Political Behavior*, 37:3–26.

Aybas, Y. C. and Turkel, E. (2022). Persuasion with coarse communication.

Bergemann, D. and Morris, S. (2019). Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95.

Bloedel, A. W. and Segal, I. R. (2018). Persuasion with rational inattention. *Available at SSRN 3164033*.

Blume, A. and Board, O. (2013). Language barriers. *Econometrica*, 81(2):781–812.

Corrao, R. (2021). Targeting in networks and markets: An information design approach. *Working Paper*.

Egorov, G. and Sonin, K. (2020). Persuasion on networks. Working Paper 27631, National Bureau of Economic Research.

Eliaz, K., Spiegler, R., and Thysen, H. C. (2021). Strategic interpretations. *Journal of Economic Theory*, 192:105192.

Galperti, S. and Perego, J. (2020). Information systems. *Working Paper*.

Hagenbach, J. and Koessler, F. (2020). Cheap talk with coarse understanding. *Games and Economic Behavior*, 124:105–121.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Le Treust, M. and Tomala, T. (2019). Persuasion with limited communication capacity. *Journal of Economic Theory*, 184:104940.

Levy, G., Barreda, I. M. d., and Razin, R. (2022). Persuasion with correlation neglect: a full manipulation result. *American Economic Review: Insights*, 4(1):123–138.

Liporace, M. (2021). Persuasion in networks. *Working Paper*.

Schwartzstein, J. and Sunderam, A. (2021). Using models to persuade. *American Economic Review*, 111(1):276–323.

Appendices

Appendix A

Appendix for Chapter 1

A.1 Proof of lemma 1

Proof. Let $\sigma \in \Sigma$ and suppose that there exist $\mu, \mu' \in \text{supp}(\sigma)$ with $p(\mu) = p(\mu')$. Consider the following market:

$$\tilde{\mu} = \frac{\sigma(\mu)}{\sigma(\mu) + \sigma(\mu')} x + \frac{\sigma(\mu')}{\sigma(\mu) + \sigma(\mu')} x'.$$

By the convexity of $X_{p(\mu)}$, $p(\tilde{\mu}) = p(\mu)$. Define σ' in the following way: $\sigma'(\tilde{\mu}) = \sigma(\mu) + \sigma(\mu')$, $\sigma'(\mu) = \sigma'(\mu') = 0$ and $\sigma' = \sigma$ otherwise. Is it easy to check that $\sum_{\mu \in \text{supp}(\sigma)} \sigma(\mu) \, W(\mu) = \sum_{\mu \in \text{supp}(\sigma')} \sigma'(\mu) \, W(\mu)$. We can iterate this operation as many times as the number of pairs $\nu, \nu' \in \text{supp}(\sigma')$ such that $p(\nu) = p(\nu')$ to finally obtain the desired conclusion. $\square$

A.2 Proof of lemma 2

Proof. Let $\mu^\star$ be an inefficient aggregate market, hence for any optimal segmentation $\sigma \in \Sigma(\mu^\star)$, $|\text{supp}(\sigma)| \geq 2$. Let σ be a direct and optimal segmentation of $\mu^\star$ and $\mu \in \text{supp}(\sigma)$ such that μ is in the interior of $X_{p(\mu)}$. Let ν be any other market in the support of σ. Consider the market:

$$\xi = \frac{\sigma(\mu)}{\sigma(\mu) + \sigma(\nu)} \mu + \frac{\sigma(\nu)}{\sigma(\mu) + \sigma(\nu)} \nu.$$

Because $\mu^\star$ is inefficient, it is without loss of generality to assume that ξ is also inefficient.

Denote $\bar{\mu}$ (resp. $\bar{\nu}$) the projection of ξ on the boundary of the simplex M in direction of μ (resp. ν). For σ to be optimal, the segmentation of ξ between μ with

probability $\frac{\sigma(\mu)}{\sigma(\mu)+\sigma(\nu)}$ and ν with probability $\frac{\sigma(\nu)}{\sigma(\mu)+\sigma(\nu)}$ must be optimal. In particular, it must be optimal among any segmentation on $[\bar{\mu}, \bar{\nu}]$.

There exists a one-to-one mapping $f\colon [\bar{\mu}, \bar{\nu}] \to [0,1]$ such that for any $\gamma \in [\bar{\mu}, \bar{\nu}]$, $\gamma = f(\gamma)\bar{\mu} + (1 - f(\gamma))\bar{\nu}$. Thus, the set $[\bar{\mu}, \bar{\nu}]$ can be seen as all the distributions on a binary set of states of the world $\{\bar{\mu}, \bar{\nu}\}$, where for any $\gamma \in [\bar{\mu}, \bar{\nu}]$, $f(\gamma)$ is the probability of $\bar{\mu}$.

Therefore, the maximization program,

$$\max_{\sigma} \sum_{\gamma \in \mathrm{supp}(\sigma)} \sigma(\gamma) W(\gamma) \tag{$\bar{S}$}$$

$$\text{s.t. } \sigma \in \Sigma^{[\bar{\mu},\bar{\nu}]}(\xi) \equiv \left\{ \sigma \in \Delta([\bar{\mu}, \bar{\nu}]) \;\middle|\; \sum_{\gamma \in \mathrm{supp}(\sigma)} \sigma(\gamma)\gamma = \xi, \ \mathrm{supp}(\sigma) < \infty \right\},$$

is a bayesian persuasion problem (Kamenica and Gentzkow, 2011), with a binary state of the world and a finite number of actions. Hence, applying theorem 1 in Lipnowski and Mathevet (2017), there exists an optimal segmentation only supported on extreme points of sets $M \in \mathcal{M}^{[\bar{\mu},\bar{\nu}]} \equiv \big\{ M_k \cap [\bar{\mu}, \bar{\nu}] \mid k \in \{1, \ldots, K\} \text{ and } M_k \cap [\bar{\mu}, \bar{\nu}] \neq \emptyset \big\}$. It happens that for any $M \in \mathcal{M}^{[\bar{\mu},\bar{\nu}]}$, so that $M = M_k \cap [\bar{\mu}, \bar{\nu}]$ for some k, if γ is an extreme point of M, then it is on the boundary of (M_k).

Let (μ', ν') with respective probabilities $(\alpha, 1 - \alpha)$ be a solution to $(\bar{S})$ where μ' and ν' are extreme points of some $M \in \mathcal{M}^{[\bar{\mu},\bar{\nu}]}$. We now consider the segmentation $\bar{\sigma}$ such that $\bar{\sigma}(\gamma) = \sigma(\gamma)$ for all $\gamma \in \mathrm{supp}(\sigma) \setminus \{\mu, \nu\}$, $\bar{\sigma}(\mu') = (\sigma(\mu) + \sigma(\nu))\alpha$, $\bar{\sigma}(\nu') = (\sigma(\mu) + \sigma(\nu))(1 - \alpha)$, and $\bar{\sigma} = 0$ otherwise. One can easily check that $\bar{\sigma} \in \Sigma(\mu^\star)$. If $\bar{\sigma}$ is not direct, that is, there exists $\gamma \in \mathrm{supp}(\bar{\sigma})$ such that (w.l.o.g.) $p(\gamma) = p(\mu')$, then construct a direct segmentation $\bar{\bar{\sigma}}$ following the same process as in the proof of lemma 1. Then, if $\bar{\bar{\sigma}}$ is not only supported on boundaries of sets $\{M_k\}_{k \in I(\mu^\star)}$, reiterate the same process as above, until you reach the desired conclusion. □

A.3 Proof of proposition 2

Proof. Fix an aggregate market $\mu^\star$ and let $\sigma \in \Sigma(\mu^\star)$ be optimal and direct. Suppose by contradiction that there exist $\mu, \mu' \in \mathrm{supp}(\sigma)$ such that $v_a := \min\{\mathrm{supp}(\mu)\} < \max\{\mathrm{supp}(\mu')\} =: v_d$ and $v_b := \min\{\mathrm{supp}(\mu')\} < \max\{\mathrm{supp}(\mu)\} =: v_c$. Assume further, without loss of generality, that $\min\{\mathrm{supp}(\mu)\} < \min\{\mathrm{supp}(\mu')\}$.

Define $\bar{\mu} := \frac{\sigma(\mu)}{\sigma(\mu)+\sigma(\mu')}\mu + \frac{\sigma(\mu')}{\sigma(\mu)+\sigma(\mu')}\mu'$. A consequence of σ being optimal is that $V(\bar{\mu}) = \frac{\sigma(\mu)}{\sigma(\mu)+\sigma(\mu')}W(\mu) + \frac{\sigma(\mu')}{\sigma(\mu)+\sigma(\mu')}W(\mu')$. The proof consists in showing that we can improve on this splitting of $\bar{\mu}$ and thus obtains a contradiction.

Define, for small $\epsilon > 0$, $\breve{\mu}, \breve{\mu}'$ as follows:

$$\breve{\mu}_k = \begin{cases} \mu_k + \epsilon & \text{if } k = b \\ \mu_k - \epsilon & \text{if } k = c \\ \mu_k & \text{otherwise.} \end{cases}$$

$$\breve{\mu}'_k = \begin{cases} \mu'_k - \frac{\sigma(\mu)}{\sigma(\mu')}\epsilon & \text{if } k = b \\ \mu'_k + \frac{\sigma(\mu)}{\sigma(\mu')}\epsilon & \text{if } k = c \\ \mu'_k & \text{otherwise.} \end{cases}$$

By construction, $\bar{\mu} = \frac{\sigma(\mu)}{\sigma(\mu)+\sigma(\mu')}\breve{\mu} + \frac{\sigma(\mu')}{\sigma(\mu)+\sigma(\mu')}\breve{\mu}'$. Note that v_a is still an optimal price for $\breve{\mu}$. Indeed, for any $v_a \leq v_k \leq v_b$, the profit made by fixing price v_k is equal in markets μ and $\breve{\mu}$ and for any $v_b < v_k \leq v_c$ the profit made by fixing price v_k is strictly lower in $\breve{\mu}$ than in μ. On the contrary, $\phi(\breve{\mu}') \geq \phi(\mu')$ and it is possible that the inequality holds strictly. In any case, it must be that $\phi(\breve{\mu}') = v_e$ for $b \leq e \leq d$. Denote $\alpha := \frac{\sigma(\mu)}{\sigma(\mu)+\sigma(\mu')}$, hence $\frac{\sigma(\mu)}{\sigma(\mu')} = \frac{\alpha}{1-\alpha}$.

$$\alpha W(\breve{\mu}) + (1-\alpha)W(\breve{\mu}') - \big(\alpha W(\mu) + (1-\alpha)W(\mu')\big) \tag{A.1}$$

$$=\alpha\big(W(\breve{\mu}) - W(\mu)\big) + (1-\alpha)\big(W(\breve{\mu}') - W(\mu')\big) \tag{A.2}$$

$$=\alpha\epsilon\big(\lambda_b(v_b - v_a) - \lambda_c(v_c - v_a)\big) \tag{A.3}$$

$$+ (1-\alpha)\Big(-\sum_{k>e}\lambda_k\mu'_k(v_e - v_b) - \sum_{b<k\leq e}\lambda_k\mu'_k(v_k - v_b) + \lambda_c\frac{\alpha}{1-\alpha}\epsilon(v_c - v_e)\Big) \tag{A.4}$$

$$=\alpha\epsilon\lambda_b(v_b - v_a) - \alpha\epsilon\lambda_c(v_e - v_a) - (1-\alpha)\Big(\sum_{k>e}\lambda_k\mu'_k(v_e - v_b) + \sum_{b<k\leq e}\lambda_k\mu'_k(v_k - v_b)\Big) \tag{A.5}$$

$$>\alpha\epsilon\lambda_b(v_b - v_a) - \alpha\epsilon\lambda_{b+1}(v_e - v_a) - (1-\alpha)\Big(\sum_{k>e}\lambda_{b+1}\mu'_k(v_e - v_b) + \sum_{b<k\leq e}\lambda_{b+1}\mu'_k(v_k - v_b)\Big) \tag{A.6}$$

$$=\alpha\epsilon\lambda_b(v_b - v_a) - \lambda_{b+1}\Big[\alpha\epsilon(v_e - v_a) - (1-\alpha)\Big(\sum_{k>e}\mu'_k(v_e - v_b) + \sum_{b<k\leq e}\mu'_k(v_k - v_b)\Big)\Big] \tag{A.7}$$

Finally,

$$(A.7) \geq 0 \iff \frac{\lambda_b}{\lambda_{b+1}} \geq \kappa$$

where

$$\kappa = \frac{\alpha\epsilon(v_e - v_a) - (1-\alpha)\big(\sum_{k>e}\mu'_k(v_e - v_b) + \sum_{b<k\leq e}\mu'_k(v_k - v_b)\big)}{\alpha\epsilon(v_b - v_a)}$$

which ends the proof. □

A.4 Proof of proposition 3

Proof. As argued in the core of the text, all markets with uniform price v_u belonging to no-rent region must be optimally segmented by splitting μ^* between $\mu^s = (\frac{\mu_1^*}{\sigma}, \frac{\mu_2^*}{\sigma}, \dots, \mu_u^s, 0, \dots, 0)$ and $\mu^r = (0, 0, \dots, \mu_u^r, \frac{\mu_{u+1}^*}{1-\sigma}, \dots, \frac{\mu_K^*}{1-\sigma})$. Such a segmentation indeed gives no rents to the monopolist if v_u is an optimal price in both μ^s and μ^r. That is, if:

$$v_1 = v_u\mu_u^s \geq v_j(\sum_{i=j}^{u-1}\frac{\mu_i^*}{\sigma} + \mu_u^s) \qquad \forall\, 2 \leq j \leq u-1 \tag{NR-s}$$

$$v_u \geq v_j(\sum_{i=j}^{K}\frac{\mu_i^*}{1-\sigma}) \qquad \forall\, u+1 \leq j \leq K \tag{NR-r}$$

As such, any optimal segmentation under strong redistributive preferences that maximizes consumer surplus must have $\mu_u^s = \frac{v_1}{v_u}$, $\sigma = \frac{v_u}{v_u - v_1}\sum_{i=1}^{u-1}\mu_i^*$ and $\mu_u^r = \frac{\mu_u^* v_u - \sum_{i=1}^{u}\mu_i^* v_1}{\sum_{i=u}^{K}\mu_i^* v_u - v_1}$, which pins down segmentation σ^{NR}. Conditions (NR-s) and (NR-r) are satisfied whenever σ^{NR} is efficient, which concludes the proof.

It is also interesting to note that conditions (NR-s) and (NR-r) define the no-rent region inside M_u as a convex polytope. Indeed, we can rearrange both conditions and get:

$$0 \geq -\alpha(j)\sum_{i=1}^{j-1}\mu_i^* + (1-\alpha(j))\sum_{i=j}^{u-1}\mu_i^* \qquad \forall\, 2 \leq j \leq u-1 \tag{NR-s}$$

$$-\frac{v_1}{v_j(v_u - v_1)} \geq -\beta(j)\sum_{i=u}^{j-1}\mu_i^* + (1-\beta(j))\sum_{i=j}^{K}\mu_i^* \qquad \forall\, u+1 \leq j \leq K \tag{NR-r}$$

for $\alpha(j) = \frac{v_1(v_u - v_j)}{v_j(v_u - v_1)}$ and $\beta(j) = \frac{v_u^2}{v_j(v_u - v_1)}$.

The conditions expressed above define $K-2$ half-spaces in $\mathbb{R}^K$. The no-rent region in M_u is thus given by the closed polytope defined by the intersection of such half-spaces. We can represent such polytope as follows:

$$NRR_u = \{\mu \in M_u : A\mu \leq z\},$$

with

$$A = \begin{bmatrix} S & O_S \\ O_R & R \end{bmatrix} \in \mathbb{R}^{K-2\times K} \text{ and } z = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -\frac{v_1}{v_{u+1}(v_u - v_1)} \\ \vdots \\ -\frac{v_1}{v_K(v_u - v_1)} \end{bmatrix} \in \mathbb{R}^{K-2}$$

where O_S and O_R are null matrices with, respectively, dimensions $(u-2)\times(u-1)$ and $(K-u)\times(K+1-u)$, and

$$S = \begin{bmatrix} -\alpha(2) & 1-\alpha(2) & \cdots & 1-\alpha(2) & 1-\alpha(2) \\ -\alpha(3) & -\alpha(3) & \cdots & 1-\alpha(3) & 1-\alpha(3) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\alpha(u-2) & -\alpha(u-2) & \cdots & 1-\alpha(u-2) & 1-\alpha(u-2) \\ -\alpha(u-1) & -\alpha(u-1) & \cdots & -\alpha(u-1) & 1-\alpha(u-1) \end{bmatrix} \in \mathbb{R}^{(u-2)\times(u-1)},$$

$$\text{R}= \begin{bmatrix} -\beta(u+1) & 1-\beta(u+1) & \cdots & 1-\beta(u+1) & 1-\beta(u+1) \\ -\beta(u+2) & -\beta(u+2) & \cdots & 1-\beta(u+2) & 1-\beta(u+2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\beta(K-1) & -\beta(K-1) & \cdots & 1-\beta(K-1) & 1-\beta(K-1) \\ -\beta(K) & -\beta(K) & \cdots & -\beta(K) & 1-\beta(K) \end{bmatrix} \in \mathbb{R}^{(K-u)\times(K+1-u)}$$

for $\alpha(j) = \frac{v_1(v_u - v_j)}{v_j(v_u - v_1)}$ and $\beta(j) = \frac{v_u^2}{v_j(v_u - v_1)}$.

□

Appendix B

Appendix for Chapter 2

B.1 Proof of proposition 4

Let Θ be any Polish space and let $\Delta(\Theta)$ be the set of probability measures on Θ endowed with its Borel σ-algebra, let also $\mathcal{C}_b(\Theta)$ be the set of bounded continuous and Borel-measurable real-valued functions on Θ.

For any $\eta, \mu \in \Delta(\Theta)$, by application of the Donsker-Varadhan variational formula (see Dupuis and Ellis, 1997, Lemma 1.4.3) we have

$$C(\eta, \mu) = \sup_{u(a,\cdot) \in \mathcal{C}_b(\Theta)} \int_\Theta \rho u(a, \theta) \, \eta(\mathrm{d}\theta) - \ln \left(\int_\Theta \exp\left(\rho u(a, \theta)\right) \, \mu(\mathrm{d}\theta) \right). \tag{B.1}$$

Taking the Legendre-Fenchel's dual to the variational equality (B.1) (see Dupuis and Ellis, 1997, Proposition 1.4.2) we get

$$\ln \left(\int_\Theta \exp\left(\rho u(a, \theta)\right) \, \mu(\mathrm{d}\theta) \right) = \sup_{\eta \in \Delta(\Theta)} \int_\Theta \rho u(a, \theta) \, \eta(\mathrm{d}\theta) - C(\eta, \mu). \tag{B.2}$$

Hence, we have

$$\Psi_a(\mu) = \frac{1}{\rho} \ln \left(\int_\Theta \exp\left(\rho u(a, \theta)\right) \, \mu(\mathrm{d}\theta) \right),$$

for any $a \in A$, any $\mu \in \Delta(\Theta)$ and any $\rho \in \mathbb{R}_+^*$. Moreover, the supremum in equation (B.2) is attained uniquely by the probability measure $\eta_a(\mu) \in \Delta(\Theta)$ defined by

$$\eta_a(\mu)(\tilde{\Theta}) = \frac{\int_{\tilde{\Theta}} \exp\left(\rho u(a, \theta)\right) \, \mu(\mathrm{d}\theta)}{\int_\Theta \exp\left(\rho u(a, \theta)\right) \, \mu(\mathrm{d}\theta)},$$

for any Borel set $\tilde{\Theta}$ (see, again, Dupuis and Ellis, 1997, Proposition 1.4.2).

In fact, we can extend the result beyond the case of the Kullback-Leibler diver-

gence. Define the φ-divergence between η and μ as

$$\mathrm{D}_{\varphi}(\eta||\mu) = \int_{\Theta} \varphi\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta)\right) \mu(\mathrm{d}\theta),$$

where $\varphi\colon \mathbb{R} \to \mathbb{R}_{+}$ is a proper, closed, convex and essentially smooth function such that $\varphi(1) = 0$ and such that its domain is an interval with endpoints $a < 1 < b$ (which may be finite or infinite). Let us also define the Legendre-Fenchel conjugate of φ, denoted φ^{*}, by

$$\varphi^{*}(y) = \max_{x \in \mathbb{R}} \; xy - \varphi(x)$$

for any $y \in \mathbb{R}$. Then, the following proposition holds.

Proposition 11. *Receiver's belief motivated by action a under posterior μ uniquely satisfies*

$$\varphi'\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta)\right) = \rho u(a,\theta),$$

for any $\theta \in \Theta$, any $a \in A$ and any $\mu \in \Delta(\Theta)$, while Receiver's optimal psychological payoff equals

$$\Psi_{a}(\mu) = \frac{1}{\rho} \int_{\Theta} \varphi^{*}\left(\rho u(a,\theta)\right) \mu(\mathrm{d}\theta),$$

for any $a \in A$ and any $\mu \in \Delta(\Theta)$.

Proof. This proposition is a direct application of Theorem 4.4 in Broniatowski and Keziou (2006). □

B.2 Overoptimism about preferred outcomes

Fix an $a \in A$ and let Θ_{a} be the (measurable) set of states such that $\Theta_{a} = \arg\max_{\theta \in \Theta} u(a,\theta)$. Define $\delta(a,\theta) = u(a,\theta) - u(a,\theta^{*})$ for all θ and some $\theta^{*} \in \Theta_{a}$. Remark that $\eta_{a}(\mu)(\Theta_{a})$ can be expressed as follows:

$$\begin{aligned}
\eta_{a}(\mu)(\Theta_{a}) &= \frac{\displaystyle\int_{\Theta_{a}} \exp\left(\rho u(a,\theta)\right) \mu(\mathrm{d}\theta)}{\displaystyle\int_{\Theta} \exp\left(\rho u(a,\theta)\right) \mu(\mathrm{d}\theta)} \\
&= \frac{\mu(\Theta_{a})}{\mu(\Theta_{a}) + \displaystyle\int_{\Theta \setminus \Theta_{a}} \exp(\rho\delta(a,\theta))\, \mu(\mathrm{d}\theta)}.
\end{aligned}$$

Let's define the function

$$h(\rho) = \frac{\mu(\Theta_a)}{\mu(\Theta_a) + \displaystyle\int_{\Theta\setminus\Theta_a} \exp\left(\rho\delta(a,\theta)\right)\, \mu(\mathrm{d}\theta)}$$

for any $\rho \in \mathbb{R}_+^*$.

First, remark that $h(0) = \mu(\Theta_a)$. Moreover, by Leibniz integral rule, we have

$$h'(\rho) = \frac{-\mu(\Theta_a)}{\displaystyle\int_{\Theta\setminus\Theta_a} \delta(a,\theta) \exp\left(\rho\delta(a,\theta)\right)\, \mu(\mathrm{d}\theta)} \geq 0$$

for any $\rho \in \mathbb{R}_+^*$, since $\delta(a,\theta) \leq 0$. Finally, we also have that $\lim_{\rho\to+\infty} h(\rho) = 1$. Hence the probability of payoff maximizing states is bounded below by the Bayesian posterior $\mu(\Theta_a)$, is always increasing and is converging to 1 from below. Hence, a wishful Receiver always puts more probability mass on Θ_a than a Bayesian and eventually believes that the state belongs to Θ_a with probability 1 when ρ becomes large.

B.3 Proof of lemma 3

Let us study the properties of the belief threshold μ^W as a function of ρ and payoffs. First of all, let us define the function

$$\mu^W(\rho) = \frac{\exp(\rho\underline{u}_0) - \exp(\rho\underline{u}_1)}{\exp(\rho\underline{u}_0) - \exp(\rho\underline{u}_1) + \exp(\rho\overline{u}_1) - \exp(\rho\overline{u}_0)}.$$

for any $\rho \in \mathbb{R}_+^*$. To avoid notational burden, we omit the superscript W in the proof. We can find the limit of $\mu(\rho)$ at 0 by applying l'Hôpital's rule

$$\begin{aligned}
\lim_{\rho\to 0} \mu(\rho) &= \lim_{\rho\to 0} \frac{\underline{u}_0 \exp(\rho\underline{u}_0) - \underline{u}_1 \exp(\rho\underline{u}_1)}{\underline{u}_0 \exp(\rho\underline{u}_0) - \underline{u}_1 \exp(\rho\underline{u}_1) + \overline{u}_1 \exp(\rho\overline{u}_1) - \overline{u}_0 \exp(\rho\overline{u}_0)} \\
&= \frac{\underline{u}_0 - \underline{u}_1}{\underline{u}_0 - \underline{u}_1 + \overline{u}_1 - \overline{u}_0} \\
&= \mu^B.
\end{aligned}$$

So, we are back to the case of a Bayesian Receiver whenever the cost of distortion becomes infinitely high. After multiplying by $\exp(-\rho\underline{u}_0)$ at the numerator and the

denominator of $\mu(\rho)$ we get

$$\mu(\rho) = \frac{1 - \exp(\rho(\underline{u}_1 - \underline{u}_0))}{1 - \exp(\rho(\underline{u}_1 - \underline{u}_0)) + \exp(\rho(\overline{u}_1 - \underline{u}_0)) - \exp(\rho(\overline{u}_0 - \underline{u}_0))}.$$

So the limit of μ^W at infinity only depends on the sign of $\overline{u}_1 - \underline{u}_0$ as, by assumption, $\underline{u}_1 - \underline{u}_0 < 0$ and $\overline{u}_0 - \underline{u}_0 < 0$. Hence, $\lim_{\rho\to+\infty} \mu(\rho) = 1$ when $\overline{u}_1 - \underline{u}_0 < 0$ and $\lim_{\rho\to+\infty} \mu(\rho) = 0$ when $\overline{u}_1 - \underline{u}_0 > 0$. Finally, in the case where $\underline{u}_0 = \overline{u}_1$ we have

$$\begin{aligned}\lim_{\rho\to+\infty} \mu(\rho) &= \lim_{\rho\to+\infty} \frac{1 - \exp(\rho(\underline{u}_1 - \underline{u}_0))}{2 - \exp(\rho(\underline{u}_1 - \underline{u}_0)) - \exp(\rho(\overline{u}_0 - \underline{u}_0))} \\ &= \frac{1}{2}.\end{aligned}$$

Let us now check the variations of the function. After differentiating with respect to ρ and rearranging terms, one can remark that the derivative of $\mu(\rho)$ must verify the following logistic differential equation with varying coefficient

$$\mu'(\rho) = \alpha(\rho)\mu(\rho)(1 - \mu(\rho)),$$

where

$$\alpha(\rho) = \frac{\underline{u}_0 \exp(\rho\underline{u}_0) - \underline{u}_1 \exp(\rho\underline{u}_1)}{\exp(\rho\underline{u}_0) - \exp(\rho\underline{u}_1)} - \frac{\overline{u}_1 \exp(\rho\overline{u}_1) - \overline{u}_0 \exp(\rho\overline{u}_0)}{\exp(\rho\overline{u}_1) - \exp(\rho\overline{u}_0)},$$

for all $\rho \in \mathbb{R}_+^*$, together with the initial condition $\mu(0) = \mu^B$. Hence, α completely dictates the variations of $\mu(\rho)$. Let us study the properties of the function α defined on $\mathbb{R}_+^*$. First, still applying again l'Hôpital's rule, its limits are given by

$$\begin{aligned}\lim_{\rho\to 0} \alpha(\rho) &= \frac{\underline{u}_0 - \overline{u}_0 - (\overline{u}_1 - \underline{u}_1)}{2} \\ &= \frac{1}{2}(u_0 - u_1)\end{aligned}$$

and

$$\begin{aligned}\lim_{\rho\to+\infty} \alpha(\rho) &= \underline{u}_0 - \overline{u}_1 \\ &= u_{\max}.\end{aligned}$$

Second, after rearranging terms, its derivative is given by

$$\alpha'(\rho) = \frac{(\underline{u}_0 - \underline{u}_1)^2}{\cosh(\rho(\underline{u}_0 - \underline{u}_1)) - 1} - \frac{(\overline{u}_1 - \overline{u}_0)^2}{\cosh(\rho(\overline{u}_1 - \overline{u}_0)) - 1},$$

for any $\rho \in \mathbb{R}_+^*$, where cosh is the hyperbolic cosine function defined by

$$\cosh(x) = \frac{e^x + e^{-x}}{2},$$

for any $x \in \mathbb{R}$. Remark that the function defined by

$$f(x) = \frac{x^2}{\cosh(\rho x) - 1} \tag{B.3}$$

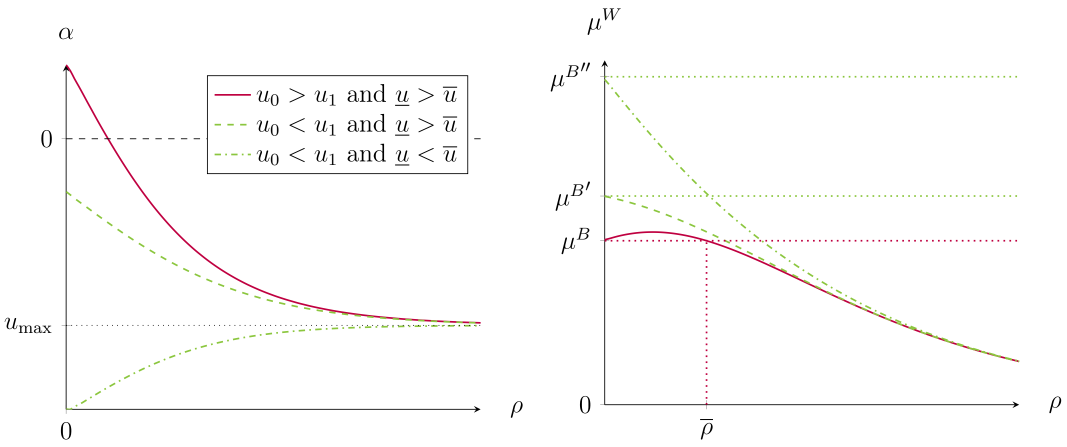
is strictly decreasing on $\mathbb{R}_+^*$. So, we have $\alpha'(\rho) < 0$ and therefore μ^W strictly decreasing for all $\rho \in \mathbb{R}_+^*$ if and only if $\underline{u}_0 - \underline{u}_1 > \overline{u}_1 - \overline{u}_0$. Accordingly, α is always a strictly monotonic function if and only if $\underline{u}_0 \neq \overline{u}_1$ and $\overline{u}_0 \neq \underline{u}_1$. Hence, excluding the extreme case where $\underline{u}_0 = \overline{u}_1$ and $\overline{u}_0 = \underline{u}_1$ so $\alpha'(\rho) = 0$ and $\mu(\rho) = \mu^B$ for all $\rho \in \mathbb{R}_+^*$, three interesting cases arise, all depicted on figure B.1 for different payoff matrices:

(i) If $u_{\max} < 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$ if and only if $u_0 < u_1$, in which case μ^W is strictly decreasing from μ^B to 0. In case $u_0 > u_1$, α has a varying sign so μ^W starts from μ^B and is sequentially strictly increasing and strictly decreasing toward 0.

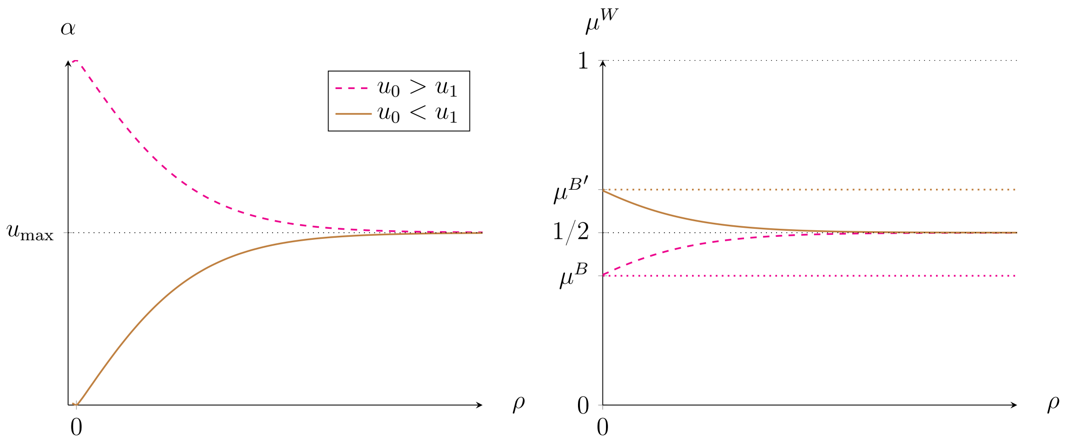
(ii) If $u_{\max} = 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$. In this case μ^W is strictly increasing from μ^B to 1/2 if and only if $u_0 > u_1$.

(iii) If $u_{\max} > 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$ if and only if $u_0 > u_1$, in which case μ^W is strictly increasing from μ^B to 1. In case $u_0 < u_1$, α has a varying sign so μ^W starts from μ^B and is sequentially strictly decreasing and strictly increasing toward 1.

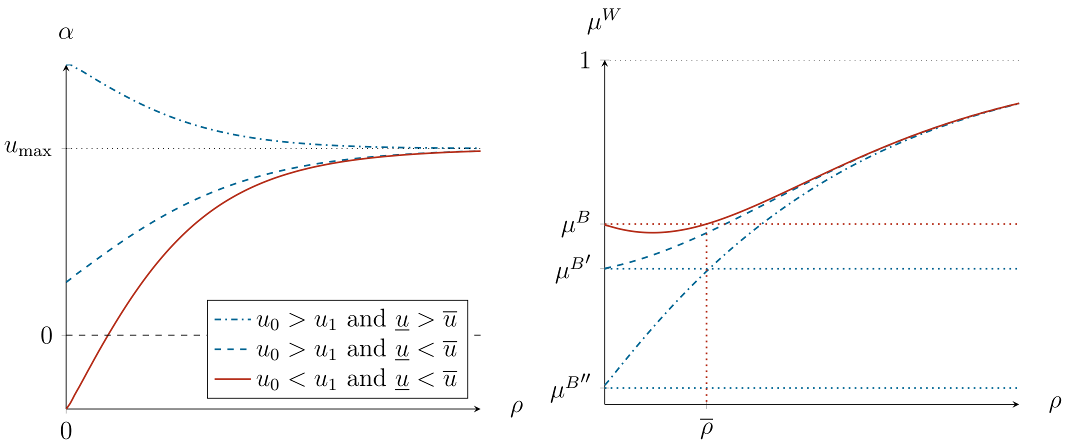
Accordingly, in case μ^W is non-monotonic in ρ, there always exists some $\overline{\rho} > 0$ such that $\mu^W(\overline{\rho}) = \mu^B$. This concludes the proof.



(a) Functions α and μ^W when $u_{\max} < 0$.



(b) Functions α and μ^W when $u_{\max} = 0$.



(c) Functions α and μ^W when $u_{\max} > 0$.

Figure B.1: Functions α and μ^W for different payoff matrices $(u_a^\theta)_{a,\theta \in A \times \Theta}$. Action $a = 1$ is favored by a wishful Receiver whenever $\mu^W < \mu^B$.

B.4 Proof of proposition 5

Assume $|\Theta| = n$ where $2 \leq n < \infty$. We want to show that $\Delta_1^B \subset \Delta_1^W$ if, and only if, the payoff matrix $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in lemma 3 for every pair of states $\theta, \theta' \in \Theta$.

Extreme point representation for Δ_1^B and Δ_1^W. First, remark that Δ_a^B and Δ_a^W are both convex polytopes in $\mathbb{R}^{|\Theta|}$ defined by

$$\Delta_a^B = \Delta(\Theta) \cap \left\{ \mu \in \mathbb{R}^{|\Theta|} : \forall a' \in A, \sum_{\theta\in\Theta} u(a,\theta)\mu(\theta) \geq \sum_{\theta\in\Theta} u(a',\theta)\mu(\theta) \right\},$$

and

$$\Delta_a^W = \Delta(\Theta) \cap \left\{ \mu \in \mathbb{R}^{|\Theta|} : \forall a' \in A, \sum_{\theta\in\Theta} \exp\left(\rho u(a,\theta)\right)\mu(\theta) \geq \sum_{\theta\in\Theta} \exp\left(\rho u(a',\theta)\right)\mu(\theta) \right\}.$$

The sets Δ_a^B and Δ_a^B are thus compact and convex sets in $\mathbb{R}^{|\Theta|}$ with finitely many extreme points. Let us now characterize the sets of extreme points of Δ_1^B and Δ_1^W. For any $\mu \in \mathbb{R}^{|\Theta|}$, define the systems of equations

$$\mathbf{A}^B \cdot \mu = \mathbf{b}, \quad \mu \geq 0$$

and

$$\mathbf{A}^W \cdot \mu = \mathbf{b}, \quad \mu \geq 0$$

where

$$\mathbf{A}^B = \begin{pmatrix} u^B(\theta_1) & \dots & u^B(\theta_n) \\ 1 & \dots & 1 \end{pmatrix},$$

and

$$\mathbf{A}^B = \begin{pmatrix} u^W(\theta_1) & \dots & u^W(\theta_n) \\ 1 & \dots & 1 \end{pmatrix},$$

are $2 \times n$ matrices, where $u^B(\theta) = u(1,\theta) - u(0,\theta)$ and $u^W(\theta) = \exp(\rho u(1,\theta)) - \exp(\rho u(0,\theta))$ for any $\theta \in \Theta$, and

$$\mathbf{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In what follows, we always assume that $(u^B(\theta))_{\theta\in\Theta}$ and $(u^W(\theta))_{\theta\in\Theta}$ are such that $\text{rank}(\mathbf{A}^B) = \text{rank}(\mathbf{A}^W) = 2$.[1] Let us recall some mathematical preliminaries.

Definition 7 (Basic feasible solution)**.** *Let $\theta, \theta' \in \Theta$ be any pair of states. A vector μ^* is a basic feasible solution to $\mathbf{A}^B \cdot \mu = \mathbf{b}$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$), $\mu \geq 0$, for θ, θ' if $\mathbf{A}^B \cdot \mu^* = \mathbf{b}$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$), $\mu^*(\theta), \mu^*(\theta') > 0$ and $\mu^*(\theta'') = 0$ for any $\theta'' \neq \theta, \theta'$.*

Lemma 5 (Extreme point representation for convex polyhedra)**.** *A vector $\mu \in \mathbb{R}^{|\Theta|}$ is an extreme point of the convex polyhedron Δ_1^B (resp. Δ_1^B) if, and only if μ is a basic feasible solution to $\mathbf{A}^B \cdot \mu = \mathbf{b}$, $\mu \geq 0$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$, $\mu \geq 0$).*

Proof. See Panik (1993) Theorem 8.4.1. □

Therefore, to find extreme points of Δ_1^B, we just have to solve the system of equations

$$\begin{cases} \mu(\theta)u^B(\theta) + \mu(\theta')b(\theta') = 0 \\ \mu(\theta) + \mu(\theta') = 1 \\ \mu(\theta), \mu(\theta') \geq 0 \end{cases} \tag{B.4}$$

for any pair of states θ, θ'. When either $\mu(\theta) = 0$ or $\mu(\theta') = 0$, the solution to (B.4) is given by the Dirac measure δ_θ only if $u^B(\theta) \geq 0$. Denote $\mathcal{E}_1^B$ the set of such beliefs. The set $\mathcal{E}_1^B$ then corresponds to the set of degenerate beliefs under which a Bayesian Receiver would take action $a = 1$. Now, if $\mu(\theta), \mu(\theta') > 0$ then the solution to (B.4) is given by

$$\mu^B_{\theta,\theta'} = \frac{u(0,\theta') - u(1,\theta')}{u(0,\theta') - u(1,\theta') + u(0,\theta) - u(1,\theta)}.$$

Such a belief is exactly the belief on the edge of the simplex between δ_θ and $\delta_{\theta'}$ at which a Bayesian decision-maker is indifferent between action $a = 0$ and $a = 1$. Denote $\mathcal{I}^B$ the set of such beliefs. Hence, we have

$$\text{ext}(\Delta_1^B) = \mathcal{E}_1^B \cup \mathcal{I}^B.$$

Following the same procedure, the set of extreme points of Δ_1^W is given by $\mathcal{E}_1^W \cup \mathcal{I}^W$, where $\mathcal{E}_1^W$ is the set of degenerate beliefs at which $u^W(\theta) \geq 0$ and $\mathcal{I}^W$ is the set of beliefs

$$\mu^W_{\theta,\theta'}(\rho) = \frac{\exp(\rho u(0,\theta')) - \exp(\rho u(1,\theta'))}{\exp(\rho u(0,\theta')) - \exp(\rho u(1,\theta')) + \exp(\rho u(0,\theta)) - \exp(\rho u(1,\theta))},$$

for any $\theta, \theta' \in \Theta$. Now, applying Krein-Milman theorem, we can state that

$$\Delta_1^B = \text{co}\left(\mathcal{E}_1^B \cup \mathcal{I}^B\right)$$

[1] This amounts to assuming that payoff are not constant across states.

and

$$\Delta_1^W = \text{co}\left(\mathcal{E}_1^W \cup \mathcal{I}^W\right)$$

Sufficiency. Assume the payoff matrix $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in lemma 3 for every pair of states $\theta, \theta' \in \Theta$. Therefore, we have $\mu^W_{\theta,\theta'}(\rho) > \mu^B_{\theta,\theta'}$ for any $\theta, \theta' \in \Theta$. This implies $\mathcal{I}_1^B \subset \Delta_1^W$, since action $a = 1$ is favored by a wishful Receiver on each edge of the simplex. Moreover, it is trivially satisfied that $\mathcal{E}_1^B = \mathcal{E}_1^W$. Hence, since any point in Δ_1^B can be written as a convex combination of points in $\mathcal{E}_1^B \cup \mathcal{I}^B \subset \Delta^W$, it follows that $\Delta_1^B \subset \Delta_1^W$.

Necessity. Assume now that $\Delta_1^B \subset \Delta_1^W$. Therefore, we have $\mu^W_{\theta,\theta'}(\rho) > \mu^B_{\theta,\theta'}$ for any $\theta, \theta' \in \Theta$ which implies that $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in lemma 3 for every pair of states $\theta, \theta' \in \Theta$.

B.5 Proof of proposition 8

First, note that we can always index the voters in an ascending order of β, such that $\eta(\mu, \beta^i) \geq \eta_j(\mu)$ for all $\mu \in \Delta(\Theta)$ whenever $i < j$, such that

$$\pi(\mu) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta(\mu, \beta^i) - \eta(\mu, \beta^j)$$

does indeed represent the absolute difference between each pair of beliefs. Now, remark that the sum can be rearranged in the following way:

$$\begin{aligned}
\pi(\mu) =& \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta(\mu, \beta^i) - \eta(\mu, \beta^j) \\
=& (n-1)\eta^1(\mu) + (n-2)\eta^2(\mu) - \eta^2(\mu) + \\
& \cdots + \frac{n-1}{2}\eta(\mu, \beta^m) - \frac{n-1}{2}\eta(\mu, \beta^m) + \cdots + \\
& \eta(\mu, \beta^{n-1}) - (n-2)\eta(\mu, \beta^{n-1}) - (n-1)\eta^n(\mu) \\
=& \sum_{i=1}^{m} (n+1-2i)(\eta(\mu, \beta^i) - \eta(\mu, \beta^{n+1-i})),
\end{aligned}$$

for any $\mu \in [0,1]$, where $m = (n+1)/2$. That is, we can express it in terms of the differences in beliefs among voters who are equidistant from the median. To see that this is true, we need to first realize that each belief appears $n-1$ times in

equation (2.6) (since each belief is paired once with each of the other $n-1$ beliefs). The beliefs of voters below the median appear more often as positive than negative (the belief of the first voter is positive in all of its pairings, the belief of the second voter is positive in all of its pairing except for the pairing with the first voter, etc.), whereas the beliefs of voters above the median are more often negative than positive. If we rearrange the terms of the sum in order to pair symmetric voters, the term $(\eta(\mu, \beta^1) - \eta_n(\mu))$ appears $n-1$ times, whereas the term $(\eta_2(\mu) - \eta(\mu, \beta^{n-1}))$ appears $n-3$ times, since out of the $n-1$ times $\eta_2(\mu)$ appears on equation (2.6), $n-2$ of them are positive and 1 is negative (the converse is true for $\eta(\mu, \beta^{n-1})$). One can continue the same reasoning for all the pairs of symmetric voters, and get to the formulation of $\pi(\mu)$ presented above. Note, also, that the belief of the median voter is summed and subtracted at the same rate, such that it does not matter in our measure of polarization.

Consider the distance between beliefs of any pair of symmetric voters $\eta(\mu, \beta^i) - \eta(\mu, \beta^{n+1-i})$ for $i \in \{1, \dots, m\}$. Given our symmetry assumption these two agents are symmetric, such that $\beta^i = 1 - \beta^{n+1-i}$. It is not difficult to show that any of those pairwise distances is maximized when agent i is distorting its belief upwards and agent $n+1-i$ is distorting its belief downwards. That is, when $\mu \in [\mu^W(\beta^i), \mu^W(\beta^{n+1-i})]$.

First, the distance between symmetric beliefs in such an interval can be rewritten as

$$\eta(\mu, \beta^i) - \eta(\mu, \beta^{n+1-i}) = \frac{\mu \exp(\rho\beta^i)}{\mu \exp(\rho\beta^i) + (1-\mu)} - \frac{\mu}{\mu + (1-\mu)\exp(\rho\beta^i)}.$$

for any $i \in \{1, \dots, m\}$ and $\mu \in [\mu^W(\beta^i), \mu^W(\beta^{n+1-i})]$.

Second, by taking the first order condition in this interval and rearranging it we get

$$\frac{\mu + (1-\mu)\exp(\rho\beta^i)}{\mu \exp(\rho\beta^i) + (1-\mu)} = 1,$$

such that the difference between symmetric beliefs is maximized uniquely at

$$\mu = \mu^W(\beta^m) = \frac{1}{2},$$

for any $i \in \{1, \dots, m\}$, $\beta^i \in \,]0, 1[$ and any $\rho \in \mathbb{R}_+^*$. Since

$$\mu^W(\beta^m) = \arg\max_{\mu \in [0,1]} \eta(\mu, \beta^i) - \eta(\mu, \beta^{n+1-i})$$

for any $i \in \{1, \ldots, m\}$, we get

$$\mu^W(\beta^m) = \arg\max_{\mu \in [0,1]} \pi(\mu),$$

which concludes the proof.

B.6 Proof of proposition 7

First, we define the function

$$\psi(z) = \frac{1}{1 - F(z)} \int_z^{\overline{\theta}} \exp(\rho\theta) f(\theta)\, \mathrm{d}\theta,$$

for any $z \in [\underline{\theta}, \overline{\theta}[$ and adopt the convention that $\psi(\overline{\theta}) = \exp(\rho\overline{\theta})$. It is not difficult to show that ψ is a continuous and strictly increasing function from $\psi(\underline{\theta}) = \hat{x} < 1$ to $\psi(\overline{\theta}) = \exp(\rho\overline{\theta})$. Define similarly the function

$$\varphi(z) = \frac{1}{1 - F(z)} \int_z^{\overline{\theta}} \theta f(\theta)\, \mathrm{d}\theta,$$

for any $z \in [\underline{\theta}, \overline{\theta}[$ and $\varphi(\overline{\theta}) = \overline{\theta}$. Again, it is not difficult to show that φ is a continuous and strictly increasing function from $\varphi(\underline{\theta}) = \hat{m} < 0$ to $\varphi(\overline{\theta}) = \overline{\theta}$.

Since ψ is strictly increasing, it thus suffices to show that $\psi(\theta^B) > 1 = \psi(\theta^W)$ to prove that $\theta^W < \theta^B$. Applying Jensen's inequality, it follows that

$$\psi(z) > \exp(\rho\varphi(z)),$$

for any $z \in]\underline{\theta}, \overline{\theta}[$, where the strict inequality comes from the strict convexity of $z \mapsto \exp(\rho z)$ and the non degeneracy of F. In particular, Jensen's inequality holds with equality at $\underline{\theta}$ and $\overline{\theta}$, but, by the intermediate value theorem, it must be that θ^B (as well as θ^W) lie in the open interval $]\underline{\theta}, 0[$. Thus, we have

$$\psi(\theta^B) > 1,$$

since $\varphi(\theta^B) = 0$ and $\theta^B \neq \underline{\theta}, \overline{\theta}$.

Appendix C

Appendix for Chapter 3

C.1 Proof of lemma 4

Proof. We can denote the posterior belief of receiver i after observing a message $m \in p_i(m)$ as:

$$\mu_i(\omega|p_i(m)) = \frac{\sigma(p_i(m)|\omega)}{\sigma(p_i(m))}\mu_0(\omega) = \sum_{p_j:p_j\subseteq p_i(m)} \frac{\sigma(p_j|\omega)}{\sigma(p_i(m))}\mu_0(\omega)$$

since the set $\{p_j|p_j \subseteq p_i(m)\}$ must completely cover the set $p_i(m)$ when P_j is a refinement of P_i. Since $\mu_j(\omega|p_j)\sigma(p_j) = \sigma(p_j|\omega)\mu_0(\omega)$, it follows that:

$$\begin{aligned}
\mu_i(\omega|p_i(m)) &= \sum_{p_j:p_j\subseteq p_i(m)} \frac{\sigma(p_j)}{\sigma(p_i(m))}\mu_j(\omega|p_j) \\
&= \sum_{p_j:p_j\subseteq p_i(m)} \frac{\sigma(p_j \cap p_i(m))}{\sigma(p_i(m))}\mu_j(\omega|p_j) \\
&= \sum_{p_j:p_j\subseteq p_i(m)} \sigma(p_j|p_i(m))\mu_j(\omega|p_j)
\end{aligned}$$

□

C.2 Proof of proposition 10

Proof. From proposition 9 we know that the distribution of μ_i impacts sender's payoffs not only by defining the distribution of actions of agent i, but also by constraining the distribution of any μ_j for $j > i$. One then needs to "backwards induct" on the beliefs in order to determine the proper value of μ_i.

From corollary 2 of Kamenica and Gentzkow (2011) and proposition 9 of this paper we know that given any μ_{n-1}, the value of receiver n's subtext for the sender

must be given by $\sup\{z|(\mu_{n-1}, z) \in \text{co}(\hat{v}_n)\} \equiv V_n(\mu_{n-1})$, such that the value of inducing a particular belief μ_{n-1} should be $\hat{v}_{n-1}(\mu_{n-1}) + V_n(\mu_{n-1})$.

Given that, one could compute the value of $n-1$'s subtext given any μ_{n-2} as $\sup\{z|(\mu_{n-2}, z) \in \text{co}(\hat{v}_{n-1} + V_n)\} \equiv V_{n-1}(\mu_{n-2})$. Recursing the argument until receiver $V_1(\mu_0)$ obtains the proof.

$\square$

Bibliography

Broniatowski, M. and Keziou, A. (2006). Minimization of ϕ-divergences on sets of signed measures. *Studia Scientiarum Mathematicarum Hungarica*, 43(4):403–442.

Dupuis, P. and Ellis, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Lipnowski, E. and Mathevet, L. (2017). Simplifying bayesian persuasion. *Working Paper*.

Panik, M. J. (1993). *Fundamentals of Convex Analysis*. Springer Netherlands, Dordrecht.